2020

Detection of Pathological HFO Using Supervised Machine Learning and iEEG Data

Isabel L. Sicardi Rosell
Technological University Dublin

Recommended Citation

Sicardi Rosell, I.L. (2020). Detection of Pathological HFO Using Supervised Machine Learning and iEEG Data. *A dissertation submitted in partial fulfilment of the requirements of Technological University Dublin for the degree of M.Sc. in Computing (Data Analytics).* doi:10.21427/ppgz-e055

Detection of pathological HFO using Supervised Machine Learning and iEEG data

Isabel L. Sicardi Rosell

A dissertation submitted in partial fulfilment of the requirements of
Technological University Dublin for the degree of
M.Sc. in Computing (Data Analytics)

September 2020

Declaration

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Technological University Dublin and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: Isabel L. Sicardi Rosell

Date: 28th September 2020

Abstract

Epilepsy is the second most common neurological disorder and it affects approximately 50 million people worldwide. One of the main characteristics of this disorder is the presence of recurrent seizures which tend to be controlled through medication. Nonetheless, 20% of the patients with this disorder are resistant to drug treatment meaning that they need to go through alternative procedures. One common option is the surgical resection of the epileptogenic tissue in the region of the brain that is responsible for the seizures. This study presents a supervised machine learning approach to identify pathological oscillations in the seizure onset zone based on the presence of high frequency oscillations. The method implements a SVM algorithm using features extracted from iEEG data. The model was trained, validated and tested with 25 patients suffering from refractory epilepsy and then evaluated in a dataset containing additional 8 patients. The algorithm was capable of detecting 77% of the positive cases in the test dataset and an average of 61% in the dataset of additional patients. Also, it presented an AUC of 77% and 85% accuracy in the test dataset. In addition, this work also discusses which are the features extracted from iEEG data that are more relevant to identify the pathological HFO and are able to satisfactorily represent the characteristics of the signals. In this respect, frequency domain features proved to be among the best model predictors. The results indicate that when these attributes and time-frequency domain attributes were added the model performance increased significantly.

Keywords: Epilepsy, HFO, SVM, iEEG signal, Supervised Learning

Acknowledgments

I would like to express my sincere thanks to my supervisor, Dr. Lucas Rizzo for his assistance and guidance throughout the dissertation process, and the many useful suggestions.

I would like to thank Balazs Vajna and the 4DigitalAge team for providing the data and introducing me to the topic, as well as help me to gain insights through thoughtful discussions.

Finally, all the appreciation and love to my family and Gaston that without their encouragement and unconditionally support I would have not been able to succeed during this course.

Contents

List of Figures

List of Tables

List of Acronyms

ANN	Artificial neural networks
AUC	Area under the ROC curve
CFC	Cross frequency coupling
DT	Decision tree
FR	Fast ripple
GB	Gradient boosting
HFO	High frequency oscillations
iEEG	Intracranial electroencephalographic monitoring
LFO	Low frequency oscillations
nHFO	Non-pathological HFO
NPV	Negative predictive value
PAC	Phase amplitude coupling
pHFO	Pathological HFO
PLV	Phase locking values
RF	Random forest
ROI	Region of interest
SOZ	Seizure onset zone
SVM	Support vector machine

Chapter 1

Introduction

1.1 Background

Epilepsy is a neurological chronic disease characterized by the occurrence of recurring spontaneous seizures. According to the World Health Organization, around 50 million people in the world have epilepsy (Fisher et al., 2014), making it the second most common neurological disorder after cerebrovascular disease (Moritz, Fox, Luscombe, & Kraemer, 1997).

The seizures tend to be caused by abnormal discharges of brain neurons and they are commonly controlled by medication. However, more than 20% of patients remain resistant to drug treatment (Kwan & Brodie, 2000; Picot, Baldy-Moulinier, Daurès, Dujols, & Crespel, 2008) which can contribute to an increase in the morbidity and mortality of this disease. Although the causes of why a patient is resistant to drug treatment are unknown, one alternative treatment for patients with refractory epilepsy is the surgical resection of the epileptogenic region responsible for seizures. This resective surgery has the potential to fully control and eliminate seizures in selected patients that are refractory to medications (De Tisi et al., 2011; Liu et al., 2018).

The accurate localization of the pathological region normally referred to as seizure onset zone (SOZ) or region of interest (ROI), is critical for the success of the surgery. To localize the focal epileptogenic area and the ROI, it is required an intracranial electroencephalographic monitoring (iEEG) over a large period of time and detailed

visual inspection of these data by medical experts (Liu et al., 2016).

Although resective surgery based on careful analysis of iEEG recordings have been used by epileptologists since the 1950s, epilepsy surgery is ofttimes not successful (Gliske et al., 2016). Therefore, several computer based solutions have been developed to assist neurologists in the identification of the seizure onset zone (Chen, Wan, Xiang, & Bao, 2017). The SOZ can be identified by the implementation of iEEG which would capture seizures' occurrences. From the iEEG recordings different types of biomarkers [1] can be measured, high frequency oscillations (HFO) being one of the most correlated to a successful outcome of surgical procedures. One of the reasons for this is that clinical procedures that focus on frequency bands lower than 30 Hz. are successful in only 50–60% of patients undergoing the surgery with significantly lower outcome rates for certain types of epilepsy (Gliske et al., 2016). In addition, the dynamic variation of phase amplitude coupling (PAC) has been used to study the differences between epileptogenic regions (Amiri, Frauscher, & Gotman, 2016).

Although defining the limits of the epileptogenic region is a difficult task, research suggests that the extraction of relevant features from the epileptic activity recorded on iEEG data can be a potential solution to discriminate seizure onset zones on epilepsy patients (Varatharajah et al., 2018).

1.2 Research Project

The current gold standard process to identify pathological oscillations in the region of interest is based on the analysis of interictal epileptiform activity as well as on the visual inspection of the iEEG. This analysis consists of patients staying in the hospital for several days while highly trained medical professionals label the onset times manually. Epileptologists rely on data based on long-term iEEG recordings captured using a large number of intracranial electrodes implanted into the brain through surgical procedure. This is a time consuming manual process that costs

[1] A biomarker is defined as an objectively measured characteristic of a normal or pathological biologic process (Dümpelmann, Jacobs, & Schulze-Bonhage, 2015)

medical resources and time. Moreover, there were cases where visual inspection of the iEEG recordings to identify the SOZ resulted in poor surgical outcome. Less than 50% of patients that went under surgery resulted in seizure free ten years after the procedure (Elahian, Yeasin, Mudigoudar, Wheless, & Babajani-Feremi, 2017).

Even though researchers have been studying and developing different approaches and algorithms to mitigate the workload of clinicians during this process, it is still possible to improve the current methodologies and techniques to detect SOZ. High-frequency oscillations (HFO) are local electrical signals recorded by electrodes and have received intense interest as potential biomarkers to improve the identification of the region of interest. Although most researchers have focused on HFO as an indication of pathological activity, these oscillations can occur across different brain areas which make them difficult to identify. Pathological HFOs can have similar characteristics (i.e. amplitude, duration and frequency) to physiological HFOs, therefore the use of conventional features do not seem to be enough to discriminate accurately the SOZ (Liu et al., 2018).

There are several challenges that prevent researchers from achieving a conclusive solution to this problem. Algorithms should be able to identify electrode's signatures at the moment of seizure to determine the SOZ. However, neuronal oscillations are not isolated and independent; they can interact with each other and can modulate the oscillations in other frequency bands causing what is called cross frequency coupling (CFC) (Amiri et al., 2016).

Based on the key points raised above, this study focuses on understanding which features and algorithms can best perform the task of discriminating pathological and physiological oscillations on patients with refractory epilepsy. Therefore, the research question that serves as the basis for this work is: *Are frequency and time-frequency domain features from iEEG signal data statistically significant to discriminate HFO in the seizures onset zones in epilepsy patients?*

1.3 Research Objectives

This research has two main objectives, the first being the identification of the most significant intra-patient features for the detection of spikes in the signal captured by iEEG data and the second one being the development of a statistical model using machine learning algorithms to detect HFOs and discriminate seizures onset zones in epilepsy patients. Table 1.1 presents the hypothesis related to these objectives and the methods used in order to accomplish them:

Objective 1	Identify relevant features to find HFOs on SOZ
Method 1	Feature selection based on machine learning algorithms
Hypothesis 1	Time-Frequency domain features will be better predictors of pathological HFOs
Objective 2	Investigate most effective machine learning algorithms to identify pathological HFOs
Method 2	Evaluation of algorithm performance metrics such as Accuracy, AUC and F1 Score
Hypothesis 2	Support Vector Machine algorithm will have a higher predictive power than other traditional supervised learning algorithms

Table 1.1: Objectives and Hypotheses of the study

The reason to investigate the most prominent features is to assess which characteristics of a patient's iEEG signal data are relevant to find the pathological HFOs and thus the seizure onset zone. This identification can potentially save epileptologists examination time of hours of iEEG recordings. Furthermore, the intention behind investigating the different machine learning algorithms is to provide an analysis of the most effective techniques and learn how they can be implemented in this context.

1.4 Research Methodologies

Defining the appropriate research methodology is key to successfully achieve the research objectives defined in the previous section. Thus this study can be classified in four categories: by its type, by its objective, by its form and by its reasoning.

This is a secondary research type of study, since it summarises and expands existing research. Since the investigation of the most relevant features and the SOZ detection strategy uses supervised machine learning, which requires subjective human involvement to collect the training data, this is classified as quantitative research. Therefore, identifying the pathological HFOs in signal data can be better addressed by extracting different types of features, using them to train machine learning models and then compare the feature importance in the best model.

Three types of features based on the signals from iEEG data were created and selected:

- Time domain features
- Frequency domain features
- Time-Frequency domain features

Also, to identify the spikes and select the most important features the following supervised learning algorithms were used:

- Decision Trees
- Random Forest
- Super Vector Machine
- Gradient Boosting
- Artificial Neural Networks

The performance of the different statistical models built is evaluated using metrics such as sensitivity, specificity and AUC. To assess whether there is a statistically

significant difference in the accuracy of the models, the hypothesis is tested using the McNemar nonparametric statistical test. Therefore this work falls under the empirical research category.

A deductive reasoning was used based on a top-down approach; the hypothesis was built based on a pre-existent theory and then tested using the relevant techniques. The results obtained were analysed which led to a conclusion and confirmation of the hypothesis established.

1.5 Scope and Limitations

The scope of this research is limited to the application of supervised machine learning techniques to identify HFO from the seizure onset zone using iEEG signals recorded from 33 patients diagnosed with refractory epilepsy.

This study focuses on the development of a series of models trained to identify HFO in the SOZ and understanding what are the most important features extracted from the iEEG signal. The attributes to be analysed are well defined features used by other authors but that have not been combined previously in the literature. In addition, the selection of the supervised learning algorithms is based also on what has been used in previous research in combination with the most popular classifier algorithms.

However, this work also presents a few limitations. Only one type of biomarker will be tested (high frequency oscillations) and supervised machine learning algorithms were selected which requires a target attribute to be previously defined. In this research the target attribute was built based on the visual inspection from doctors. Considering the time limitation to develop this work, features extracted from additional biomarkers were not considered in this research, although literature suggests that those might enhance the model performance.

While the current sample size utilized is far superior than the one used in previous research, a second dataset was not available. Therefore, a sub-sample of patients was selected to be used to test whether the proposed solution generalizes well to unseen data. This sub-sample of 8 patients did not take part in the training nor model building

at any moment, thus it is safe to say that it could be used to perform a fair assessment of the model. Nonetheless, if there was a bias on how the iEEG recordings were collected or how the patients were selected then it could affect the model capability to generalize.

1.6 Document Outline

This thesis is organized in five chapters which are structured as follows. Chapter 2 introduces an overview of previous research and current gaps presented in the literature. The methodology used to design the experiment and the data collection is presented in Chapter 3. This chapter also includes a detailed explanation on how the time domain and frequency domain features were created, as well as a description of the algorithms and the model evaluation metrics selected. In Chapter 4, the application of a Baseline model based on traditional features is illustrated in the first half of the chapter. Then on the second half, a set of Challenger models are presented, which are built using a combination of traditional features and the additional time-frequency domain features added. In this chapter, the results of this approach and its implications are described and discussed. Finally, this work concludes with a description of the next steps and how future work can leverage from this research in Chapter 5.

Chapter 2

Literature Review

2.1 Approaches to solve the problem

Literature review suggests that there are different approaches to try to solve this problem. Some researchers have focused on the extraction of HFO traditional features such as Duration, Amplitude, Energy and Peak ratio, and used K-means clustering and Hierarchical linear mixed model to detect the region of interest (Malinowska, Bergey, Harezlak, & Jouny, 2015). Results achieved by the authors indicate that the average HFO rate is higher for SOZ channels compared to non-SOZ channels. This is consistent with the results found by (Matsumoto et al., 2013) where pathological HFO had lower average frequency, longer average duration and a higher average spectral amplitude than physiological HFO. Malladi, Johnson, Kalamangalam, Tandon, and Aazhang (2018) proposed a novel approach of using mutual information in frequency metric to infer the cross-frequency coupling mechanisms during epileptic seizures, identifying that HFO increases in ictal periods.

2.1.1 Features

There is still discussion among researchers on how HFO are defined and identified. They have been defined as: oscillations in the frequency band 65–600 Hz (Cimbalnik, Kucewicz, & Worrell, 2016); as short events defined in the band 80–500 Hz (Liu et

al., 2018; Murphy, Von Paternos, & Santaniello, 2017; Staba, Wilson, Bragin, & Fried, 2002). They have also been divided into ripples (80-200 Hz), fast ripples (FRs, 250-500 Hz) and very high frequency oscillations (VHFOs, over 1000 Hz) (Wan, Wu, Wan, & Du, 2016).

In addition, Liu et al. (2016) studied the use of more advanced features such as spectral entropy, sub-band power ratio and frequency. The results of this study imply that these features are capable of discriminating SOZ more accurately than traditional HFO attributes. Moreover, non-linear model techniques such as Kolmogorov entropy have been demonstrated to be reliable to extract different states of normal and pathological brain pattern (Van Drongelen et al., 2003). Furthermore, other types of biomarkers have been studied such as cross-frequency coupling (CFC), which in some cases combines the features from low frequency oscillations (LFO) with HFO features (Jensen & Colgin, 2007; Guirgis, Chinvarun, Del Campo, Carlen, & Bardakjian, 2015). Also, adding features measuring the strength of coupling between the amplitude and phase based was shown to be successful for identifying SOZ (Edakawa et al., 2016). Additionally, the location of iEEG electrodes involved in a seizure onset zone has also been studied as possible features represented on the high-dimensional spatial information (Zhang, Xu, Wang, & Liang, 2010).

Feature extraction plays a key role in the detection of SOZ, and becomes important to determine the valuable features from iEEG data. Siuly and Li (2015) designed mechanism for feature extraction which is based on optimum allocation methods to acquire representative sampling points and principal component analysis to eliminate redundant EEG data information. In addition, Zhang et al. (2010) proposed an automatic patient-specific SOZ detection using incremental non-linear dimensionality reduction. This unsupervised incremental learning scheme has the potential to effectively map new data into the embedded space automatically which would lead to a significant reduction of the time spent by doctors on visual inspection tasks.

2.1.2 Techniques

In recent years researchers have studied different machine learning techniques for the automated recognition of the pathological area. Still there is no consensus on what type of algorithm is preferred to perform this task, thus Support Vector Machine (SVM) (Shoeb, 2009; Güler & Übeyli, 2007; Chen et al., 2017; Dian, Colic, Chinvarun, Carlen, & Bardakjian, 2015), Artificial Neural Networks (Gabor, Leach, & Dowla, 1996; Ghosh-Dastidar & Adeli, 2009), Bayesian classifiers (Saab & Gotman, 2005), Nearest-Neighbor classifiers (Qu & Gotman, 1997) and clustering techniques (Liu et al., 2016; Malinowska et al., 2015) are listed among the most used algorithms to detect epileptic onset zones. Table 2.1 summarizes the different machine learning techniques and algorithms used in the literature:

Paper	**Dataset Size**	**Algorithm**
Iscan, Dokur, and Demiralp (2011)	Not Available	Decision Tree
Wan et al. (2016)	Not Available	Fuzzy NN
Iscan et al. (2011) Qu and Gotman (1997) **Sciaraffa,2020**	Not Available 23 patients 18 patients	KNN
Elahian et al. (2017) Modur and Miocinovic (2015) Sciaraffa et al. (2020)	18 patients 10 patients 18 patients	Logistic Regression
Iscan et al. (2011) Saab and Gotman (2005)	Not Available 28 patients	Naive Bayes
Gabor et al. (1996) Ghosh-Dastidar and Adeli (2009) Guo, Rivero, and Pazos (2010)	22 patients Not Available Not Available	Neural Network
Sciaraffa et al. (2020)	18 patients	Random Forest
Akter et al. (2020) Chen et al. (2017)	8 patients 22 patients	SVM

Cimbalnik et al. (2019)	43 patients	
Dian et al. (2015)	6 patients	
Iscan et al. (2011)	Not Available	
"Channel-Wise Characterization of High Frequency Oscillations for Automated Identification of the Seizure Onset Zone" (2020)	5 patients	
Matsumoto et al. (2013)	5 patients	
Nicolaou and Georgiou (2012)	5 sets	
Price and TS (2008)	10 patients	
Sciaraffa et al. (2020)	18 patients	
Shoeb (2009)	23 patients	
Shoeb et al. (2004)	12 patients	
Tito, Cabrerizo, Ayala, Jayakar, and Adjouadi (2010)	14 patients	
Varatharajah et al. (2018)	82 patients	
Xiang et al. (2015)	Not Available	
Chua, Chandran, Acharya, and Lim (2008)	5 patients	GMM
Liu et al. (2016)	8 patients	
Liu et al. (2018)	13 patients	
Smart and Chen (2015)	Not Available	
Malinowska et al. (2015)	45 patients	K-means

Table 2.1: Summary of techniques used in reviewed papers

2.2 Gaps in Research

Most researchers have used a single biomarker to identify the SOZ. Considering combining features from different biomarkers could lead to more precise SOZ identification

(Varatharajah et al., 2018). Some researchers have proposed to detect spikes and analyse its relationship with SOZ (Gaspard, Alkawadri, Farooque, Goncharova, & Zaveri, 2014). However, several studies selected HFO from iEEG as a good candidate for biomarker (Frauscher et al., 2017).

Different criteria have been used to identify HFO without specifying time frequency in the oscillations nor how many of them are required to distinguish HFO from background activity. Studies do not specify their definition of HFO or how they detected it. It must be calibrated to each patient and EEG dataset (Malinowska et al., 2015; Murphy et al., 2017).

Moreover, there is evidence that different types of HFO features can be relevant to identify the SOZ:

- Several studies focused on using conventional features such as: amplitude and spectral frequency (Modur & Miocinovic, 2015). Research has been limited to investigate HFO changes in time and its relationship with spikes and SOZ, but they do not analyse HFO characteristics (Dümpelmann et al., 2015).

- Non-linear measures such as Lyapunov exponent (Zhang et al., 2010), the calculation of fuzzy entropy and short time energy of filtered signals (Wan et al., 2016) and features based on power spectrum (Chua et al., 2008; Jacobs et al., 2016) have been proposed. In addition, phase locking values (PLV) were proposed to extract the PLV features such as: PLV positive, PLV peak, PLV power (Elahian et al., 2017). However, non-linear methodologies to extract features are dependent on the sample size and features based on power spectrum can lose information about the high-order feature.

- Use of wavelet transformation: the optimal combination of four factors, mother wavelet, decomposition level, frequency band, and feature was explored in (Chen et al., 2017). Efficiently setting these factors will lead to high seizure detection accuracy with low computational cost.

In addition, most studies have not contemplated intra-patient variations nor the temporal variation of the epileptic activity. Intra-patient features which could help

to discriminate epileptic brain regions that should be considered. It is presumably that additional data would help to tackle the correct identification of epileptogenic zones and its patterns (Cimbalnik et al., 2016; Smart & Chen, 2015). One proposed option is the adoption of Point Process Models which estimate the likelihood of a HFO occurrence given that a sequence of past HFO has been observed (Sumsky & Santaniello, 2018). However, this analysis only considered ripples event, which are a subset of HFO.

Given the limited availability of data, many studies were developed with less than 10 patients (Alper et al., 2008; Dian et al., 2015). Analyses based on small cohorts of patients are not enough to pick up variations on signals and could lead to skewed results. Data must be representative of different epilepsy types and seizures should not be concentrated in a single brain region. In addition, research should have not tested the proposed approaches to a different dataset to assess how well it generalizes.

2.3 Summary

This chapter presented an overview of the different approaches taken to solve the problem of identifying pathological oscillation in the seizure onset zone in patients with epilepsy. As it was observed, many algorithms have been used over time to try to identify pathological oscillations and seizure onset zones, Support Vector Machines being by far the most popular among researchers.

When analysing the types of features used to train the different machine learning models, it was noticed that different methods have been used to extract features from the iEEG signal. From more traditional approaches using wavelet transform and deriving attributes from it to the extraction of features based on cross coupling frequencies, researchers have been working on different ways to characterise the signals from iEEG data.

In recent years, extensive research has been done in the field and researchers have studied different biomarkers and techniques to identify the signal characteristics within the seizure onset zone. Nonetheless, a few gaps were identified which include the fact

that different machine learning algorithms have been used, both supervised and unsupervised, but there is no consensus on the best technique to identify the oscillations belonging to the seizure onset zone. Also, different types of features have been used across all the studies but not many of them have been combined to capture the characteristics of the pathological oscillations.

The gaps presented define the base of this research, which involves the use of five different supervised machine learning algorithms that were trained using features from different domains: time, frequency and time-frequency domain. The literature review suggests that better model performance would be achieved if these features are combined to identify the oscillations that belong in the SOZ. Therefore, this investigation tries to answer the following research question: *Are frequency and time-frequency domain features from iEEG signal data statistically significant to discriminate HFO in the seizures onset zones in epilepsy patients?*

Chapter 3

Experiment design and methodology

This chapter describes the experiment setup, including data collection, preparation and a detail of the features created. It also includes an explanation of the machine learning algorithms selected and how the proposed solution was evaluated. In section 3.1, the hypothesis and objectives of this research are presented, as well as an overview of the tasks performed on experiment implementation. In section 3.2 describe the methods used to prepare the iEEG data and to create the features necessary for modeling. Finally, the subsequent sections 3.3 and 3.4 summarize the algorithms selected and the methods used for evaluation.

3.1 Hypothesis

How to identify the most relevant attributes to detect SOZ is a problem that can be better addressed by extracting features with different domains from the high frequency oscillations. If a combination of time-frequency domain features are used to discriminate pathological oscillations, then models that discriminate SOZ will show better performance than the models that use traditional intra-patient attributes.

H0: There is no statistically significant difference on the average accuracy of the model used to detect the SOZ when combined time-frequency features are used.

H1: There is a statistically significant difference on the average accuracy of the model used to detect the SOZ when combined time-frequency features are used.

In this research, time domain, frequency domain and time-frequency domain features extracted from electrical signals on the high frequency band were combined with the aim to accurately discriminate HFOs in the epileptic region of the brain. With the aim of answering the proposed research question (section 1.2), a quantitative research study was designed to identify the most relevant attributes to detect seizure onset zones (SOZ) by using machine learning models. Figure 3.1 provides an overview of the process followed to execute this study:

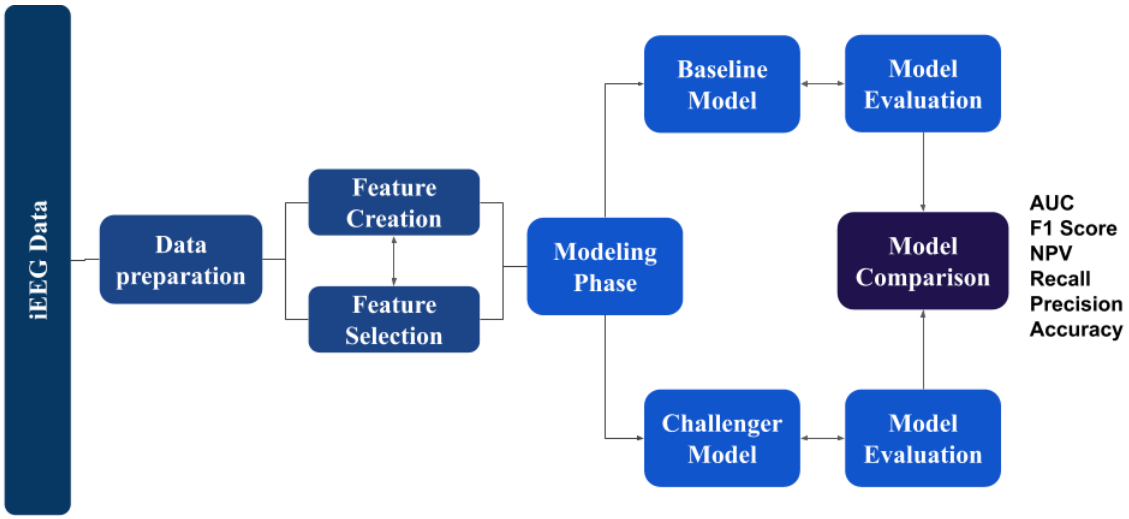

Figure 3.1: Design of the implementation for identifying SOZ

The underlying hypothesis behind the design of the present comparative experiment is that if non-traditional features extracted from iEEG data are used to discriminate pathological oscillations, then machine learning models will show better performance than the models that use only traditional attributes.

3.2 Dataset, Context and Participants

The dataset selected for developing the machine learning models contains anonymized patient information of continuous intracranial electroencephalography (iEEG) recordings of 33 patients suffering from refractory (drug-resistant) epilepsy from a Hungarian Hospital [1]. The dataset was provided in files named with the ID of the thirty three patients (e.g. ID1) and the raw data containing the recording of each patient was saved in a Neuroscan ".cnt" file. An example of the multi-channel iEEG recording for a patient is presented in Figure 3.2

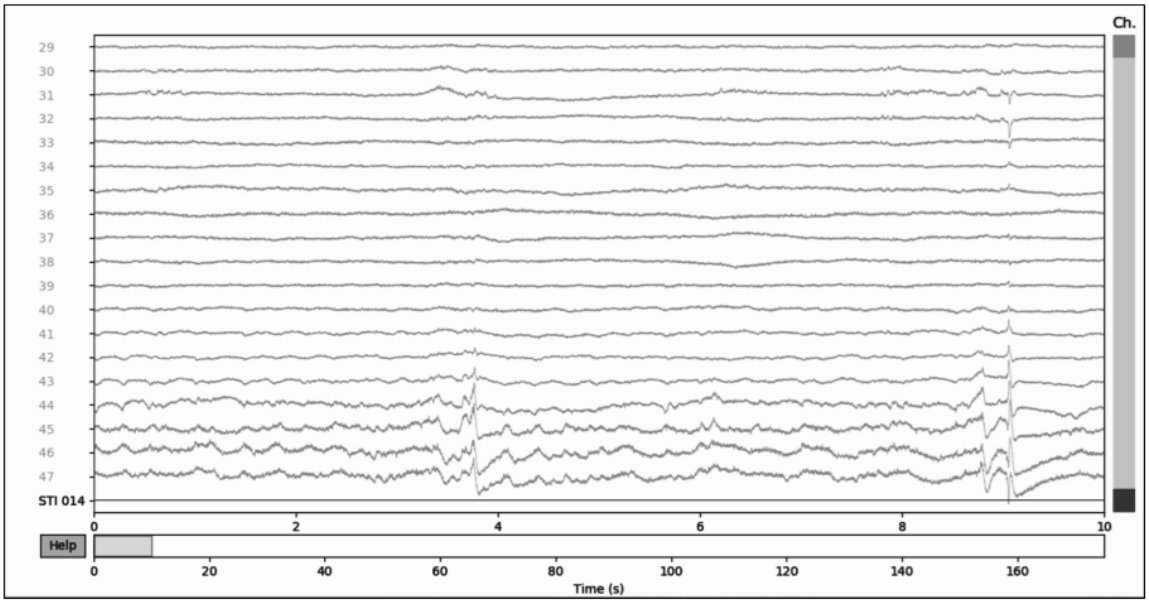

Figure 3.2: Example of 160 seconds iEEG recording for one patient

The iEEG signals were recorded intracranially by using different types of electrodes, including but not limited to strip, grid, and depth electrodes. The sampling frequency ranged between 1024 Hz and 2048 Hz and the length of the recordings range from 200 up to 2000 seconds, varying by patient. On average, 18 minutes of recordings for each subject were available and an average of 48 channels per patient were implanted.

The recording of signals took place during the night and captured pre-ictal, post-ictal or interictal stages of the different patients. All the iEEG recordings were visually

[1]The dataset was provided by a third party company.

inspected and seizure onset zones were identified by experienced epileptologists. This manual classification of the seizure onset zones was used to build the target variable used to train the models.

3.3 Methodology

The approach selected is based on the CRISP-DM methodology which provides a framework to plan a data mining project that works for different industries. This approach was selected due to the fact that is a robust and well-proven methodology used by researchers across different projects. The CRISP-DM model is presented in Figure 3.3:

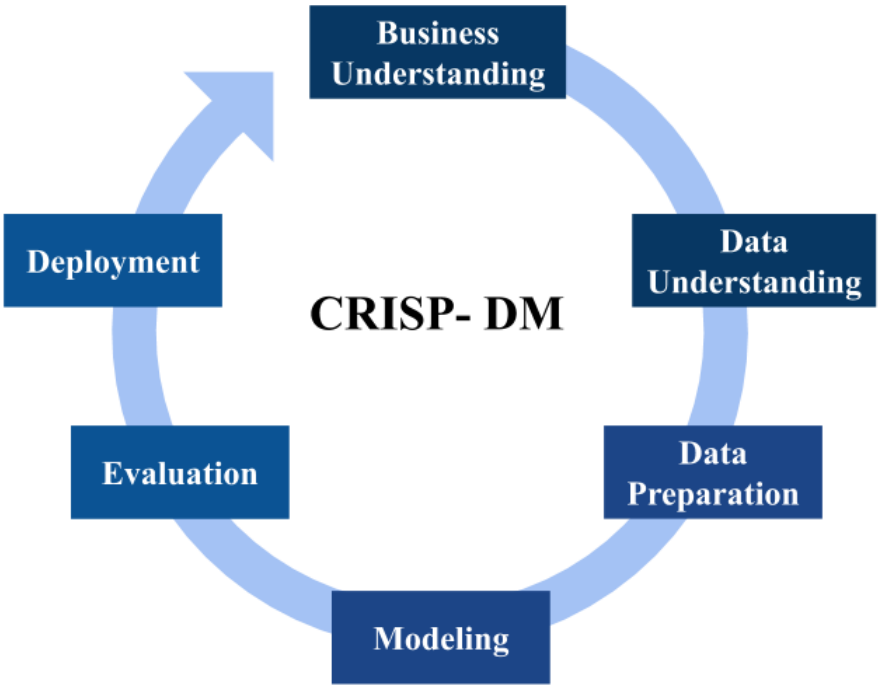

Figure 3.3: CRISP-DM Methodology, adapted from Wirth and Hipp (2000)

In line with the approach presented by the CRISP-DM methodology (Wirth & Hipp, 2000), the Data Preparation, Modeling and Evaluation stages that comprehend this experiment were implemented in five phases as presented in Figure 3.4.

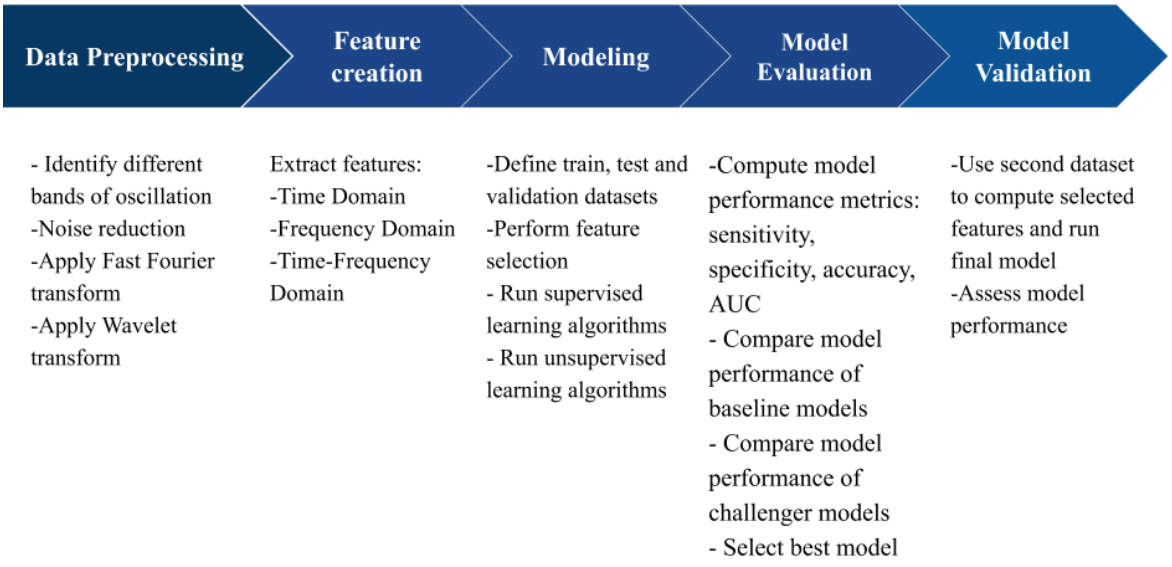

Figure 3.4: Implementation phases carried out in this study

The methods and techniques used in each phase are detailed in the subsequent subsections.

3.3.1 Data Preprocessing

The first step before the modeling phase is initiated, is the preprocessing of the data which includes analysing the signal, applying filters to remove artifacts and signal averaging. This phase consists in the definition and identification of the high frequency band and ripples events by using the algorithm proposed by (Staba et al., 2002), as well as the detection and removal of artifacts to reduce noise. The signals were digitally filtered by a finite impulse response (FIR) band-pass filter within the range 80Hz - 500Hz to remove low-frequency artifacts and Savitzky–Golay filter was applied with the purpose to increase precision without distorting the signal in the data. Figure 3.5 presents the steps taken to perform the signal processing.

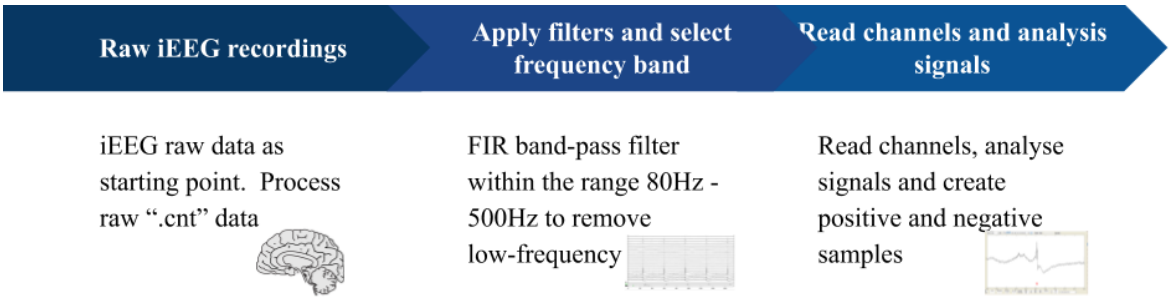

Figure 3.5: Illustrative diagram steps implemented to process the iEEG signals

3.3.2 Feature Creation

The second activity in this analysis is the feature creation phase which intends to define a vector of attributes that will be input into the models in the training phase. The goal behind extracting these features is to minimize the loss of information and resources needed to process the signal data (Al-Fahoum & Al-Fraihat, 2014).

There are several methods that have been employed in previous research to extract features from iEEG signals, among those are Wavelet transform (WT), Fourier Transform (FT) and time-frequency distributions (TFD). In this research, a Continuous Wavelet Transform and a Fourier Transform were applied to the raw signal to transform it into the required frequency domain.

Continuous Wavelet Transform (CWT): It measures the similarity between a signal and a function. Specifically, a signal is projected on a continuous family of frequency bands as expressed in equation 3.1:

$$C(a, b; f(x)), \psi(t)) = \int_{-\infty}^{\infty} f(x) \frac{1}{a} \psi^* \left(\frac{x-b}{a} \right) dx \tag{3.1}$$

Where:

- a is a scalar parameter $a > 0$
- b represents the position
- * denotes the complex conjugate
- ψ is a wavelet, representing the analyzing function

Here the Morlet Wavelet Transform was used to extract the features based on the equation 3.2:

Morlet Wavelet function

$$\psi(x) = \exp^{-\frac{x^2}{2}} \cos 5x \tag{3.2}$$

One of the most important properties of the Wavelet transform is that given any general function, it can be expressed as an infinite series of wavelets. This is particularly useful when dealing with non stationary signals such as iEEG (Al-Fahoum & Al-Fraihat, 2014).

Fast Fourier Transform (FFT): Similar to the CWT, the Fast Fourier transform also measures similarity between a signal and a function under analysis. The goal is to transform the temporal representation of a waveform $x(t)$, to a frequency domain representation $X(j\omega)$. In this case, such function is a periodic waveform represented by complex exponential, $\exp^{j\omega t}$, where t is the period. This results in a function dependent on a single variable ω as presented in equation 3.3:

$$F(\omega) = \int_{-\infty}^{\infty} f(x) \exp^{-i\omega x} \, dx \tag{3.3}$$

Using the transformations above, a total of seventeen features were extracted from iEEG data and divided into two groups: advanced time-frequency features and traditional features. Time-Frequency features provide information about the wave form and the repetitiveness of the signals in the data. On the other hand, traditional features have the advantage that they tend to be easy to compute and can give practical information about the data. These features are summarized in Table 3.1 and are described in the subsequent sections.

Time Domain Features

The main goal of the feature extraction phase is to choose the statistics and features, which are an appropriate representation of the original iEEG signal. Consecutive

values of time series tend to be highly correlated and are usually not independent, which can lead to redundancy and makes the feature extraction process more relevant (Hussein, Mohamed, Shaban, & Mohamed, 2013).

The following statistical features related to the time-domain of the signals were selected: Maximum (equation 3.4), Minimum (equation 3.5), Mean, Standard Deviation, Normalized Standard Deviation and Coefficient of Variation.

1. **Maximum (Max)**

$$maximum = \max x_i \tag{3.4}$$

2. **Minimum(Min)**

$$minimum = \min x_i \tag{3.5}$$

3. **Mean**

$$\mu(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{3.6}$$

4. **Standard Deviation (std)**

$$\sigma(x) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3.7}$$

5. **Coefficient of Variation (CV)**

$$C_v = \frac{\sigma(x)}{\mu(x)} \tag{3.8}$$

6. **Skewness**

$$sk = E\left[\frac{(x-\mu)^3}{\sigma^3}\right] \tag{3.9}$$

7. **Kurtosis**

$$k = E(x^4) - 3E(x^2)^2 \tag{3.10}$$

Where n represents the total number of data points in the window, while x_i represents the *ith* data point in the window.

Frequency Domain Features

Three statistical features were extracted for each iEEG segment:

1. **Energy**

$$energy = \sum_{i=1}^{N} (\frac{x_i(n)^2}{N}) \tag{3.11}$$

2. **Entropy**

$$entropy = -\sum_{i=1}^{N} p(x_i) log(p(x_i)) \tag{3.12}$$

3. **Root Mean Square Power (RMS)**

$$RMS = \sqrt{\frac{\sum_{i=1}^{N} |x_i^2|}{N}} \tag{3.13}$$

Time-Frequency Domain Features

Based on the transformations presented above, the following time-frequency features were extracted:

1. **Number of Peaks** If a data point is three standard deviations away from the mean, then it is classified as a peak in the signal within the observation window. Figure 3.6 illustrates how the variable was calculated:

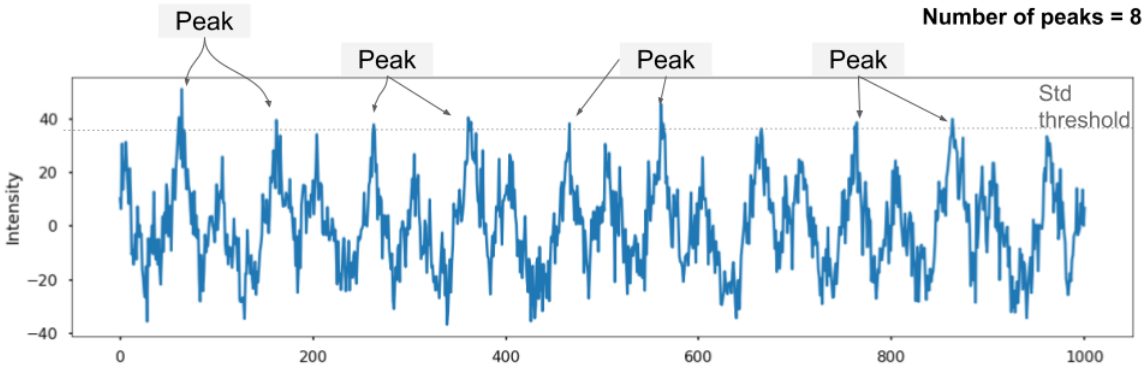

Figure 3.6: Number of peaks

2. **Average of Maximum Peak**

   Calculated as the average of the values of the identified peaks in the signal within the observation window.

3. **Peak to Notch**

$$PeaktoNotch = max\Big(\frac{Energy_{max_i}}{Energy_{min_i}}\Big) \tag{3.14}$$

Where:

(a) $Energy_{max}(x_i)$ is the energy of the maximum point of the power spectrum in the observation window

(b) $Energy_{min}(x_i)$ represents the energy of the minimum point of the power spectrum in the observation window

The figure below illustrates how this feature was created based on the description from Liu et al. (2016):

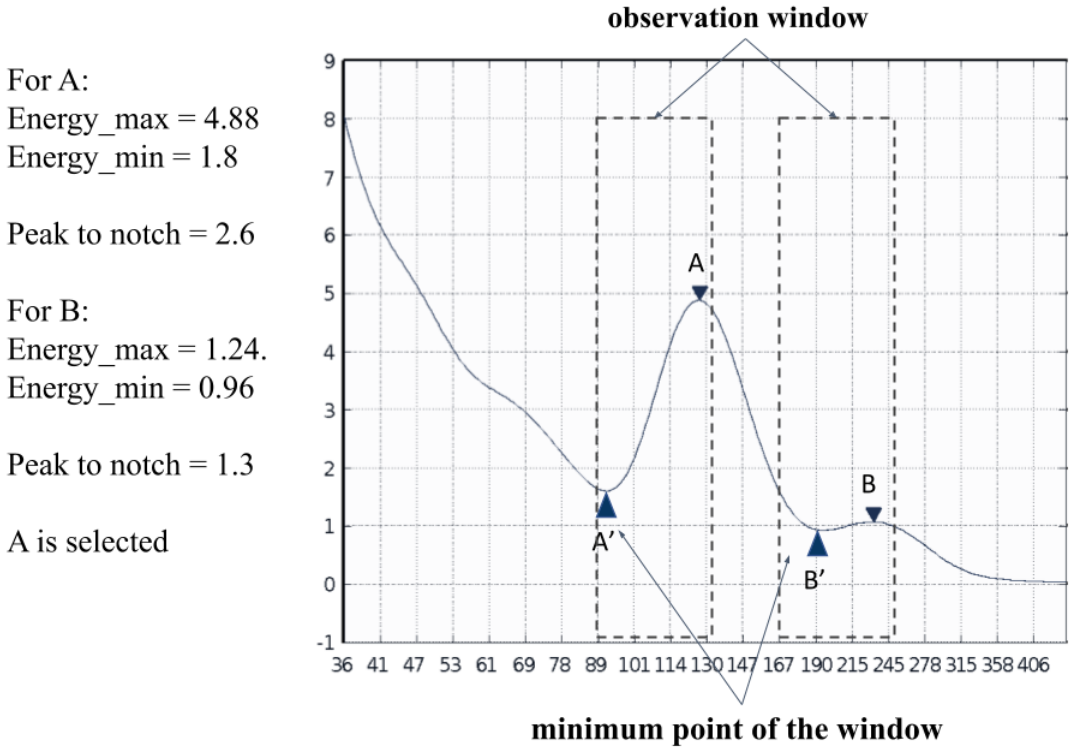

Figure 3.7: Peak to Notch

4. **Power Spectral Density (PSD)** As defined by Turitsyna and Webb (2005)

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) \exp^{-i\omega\tau}\, d\tau \tag{3.15}$$

Where:

- R_{xx} corresponds to the auto-correlation function of the signals output of a Fourier Transform

5. **Power Band Ratio (PBR)** This feature summarized the contribution of a given frequency band to the overall power of the signal.

$$PBR = \frac{S_{xx1}(\omega)}{S_{xx2}(\omega)} \tag{3.16}$$

Where:

- S_{xx1} corresponds to the power spectral density of the band frequency 1
- S_{xx2} corresponds to the power spectral density of the band frequency 2

6. **Kullback-Leibler distance**

The Kullback-Leibler distance (or relative entropy), it is a measure of how the distribution of the signal is different from a second reference probability distribution, in this case Gaussian distribution.

$$KL = \sum_i (p(x) * log\left(\frac{p(x)}{q(x)}\right) \tag{3.17}$$

Where:

- $p(x)$ corresponds to the signal distribution
- $q(x)$ corresponds to the Gaussian distribution used to approximate $p(x)$

7. **Amplitude envelope**

The amplitude envelope is given by the magnitude of the analytic signal which can be defined as:

$$x_a = F^{-}1(F(x)2U) = xi + y \tag{3.18}$$

Where:

- F is the Fourier transform
- U the unit step function
- y the Hilbert transform of x

Complete list of features

Feature	**Definition**	**Equation**
Max	Maximum value of coefficients	3.4
Min	Minimum value of coefficients	3.5
Mean	Mean value of coefficients	3.6
Std	Standard Deviation of coefficients	3.7
CV	Coefficient of variation of coefficients	3.8
Skewness	Skewness of coefficients	3.9
Kurtosis	Kurtosis of coefficients	3.10
Energy	Square sum of coefficients	3.11
Entropy	Entropy of the spectrum	3.12
RMS Power	Root mean square power	3.13
Number Peaks	Number of peaks in a window	3.14
Peak to Notch	The frequency corresponding to the maximum peak to notch ratio	3.15
Peak Maximum Avg	The average of the peaks values	3.16
Power Spectral Density	Distribution of the power into frequency components	3.17
Power Band Ratio	The sub-band power ratio after denoising the time frequency plane	3.18
KL Dist	Kullback-Leibler distance	3.19
Amplitude Envelope	Magnitude of the analytic signal	3.20

Table 3.1: Complete list of features extracted from iEEG signals

To support the decision of including these variables in the training phase, an exploratory data analysis was conducted in which the distribution of the variables and their correlation were analysed. The detailed results are presented in Appendixes A.1.1 and A.1.2. In addition, a bivariate analysis was performed to understand the relationship of the variables with the target. The results obtained are presented in Appendix A.1.3.

3.3.3 Data Preparation

After preprocessing the data and extracting the features, the following additional steps were taken to prepare the data for the modelling phase:

1. Data normalization : with the aim of avoiding bias caused by the differences in the ranges feature values, all the attributes were normalized so its distribution would have a mean of zero and a standard deviation of one ($\hookrightarrow \mathcal{N}(0,1)$) by using standardization process:

$$z = \frac{x - \mu}{\sigma} \tag{3.19}$$

   Where μ is the mean of the population and σ is the standard deviation of the population

2. Univariate Analysis: it refers to the analysis of the descriptive statistics and the features frequency using histograms.

3. Bivariate analysis: it takes into consideration each variable independently and analyses the relationship between the feature values and the dependent variable. Those variables which do not present a clear relationship with the target will be excluded from the model.

4. Data split: following common practice in the field, the dataset was split into 3 subsets; training (60%), validation(20%) and test (20%). The model coefficients are estimated from the training dataset, while the hyperparameters of the chosen

algorithm are tuned using the validation dataset. Finally, the unbiased evaluation of the performance metrics is obtained from applying the fitted model to the test dataset.

3.3.4 Modeling Phase

In this section, an overview of the most common algorithms used in the literature is given. Several existing machine learning algorithms have been described and compared among researchers, however after a detailed literature review was executed, it remains unclear which machine learning algorithm is more appropriate to address the presented problem. Therefore, the following question raises: *Which method best suits this specific application?*

The approach selected to carry out the activities of the modeling phase are described below:

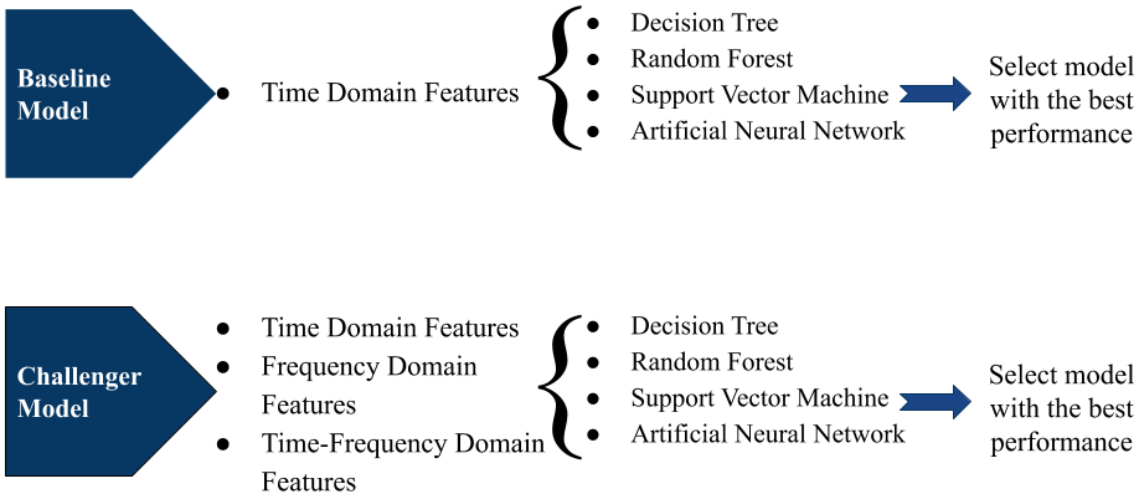

Figure 3.8: Modeling phase approach

In this study a set of different algorithms were explored and implemented to train models able to identify HFO in the epileptic region of the brain. First, a set of Baseline models were trained using time domain features and implementing the five different algorithms described in this section. From these set of model a champion was selected

based on the performance metrics described in the following section. Second, a set of Challenger models were trained adding the frequency and time-frequency domain attributes and implementing the same set of algorithms. A champion model was selected using the same approach as in the Baseline models.

In the next subsections, the supervised machine learning algorithms used are described, exploring these techniques can provide an understanding of which is the most suitable given the data and problem in question.

Decision Trees (DT)

Decision Trees are a widely-used algorithm for supervised learning problems. Its popularity is attributed to the fact that this is a simple algorithm to understand which can achieve high levels of accuracy. There are several Decision Tree algorithms (e.g. C4.5, CHAID, CART, etc.), the one selected in this study was CART algorithm implemented on sklearn library in python. This algorithm can take both categorical and numeric inputs, as well as it can handle missing values. A decision tree is illustrated in Figure 3.9

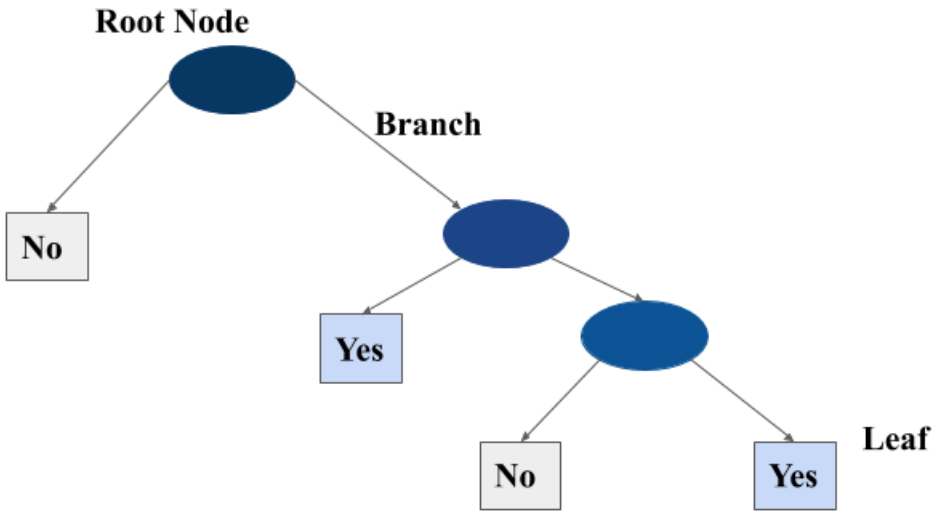

Figure 3.9: Illustrative example of Decision Tree

In this case, the feature that has the highest information gain is selected as the root node (the highest decision node). The same strategy is replicated in each subdivision of

the training dataset until all instances are divided (Hastie, Trevor Tibshirani, Robert Friedman, 2009; Kubat, 2017)

One of the many advantages of the decision tree algorithms is that their logic and output are easy to understand, even people that are not experts on the area. Although accuracy metrics are important to assess an algorithm suitability, its interpretability is equally important and that is why it has been selected to be tested in this analysis.

Random Forest (RF)

Random forests is an assemble method based on the average of a large number of randomized decision trees built by the algorithm. In recent years, this algorithm has become very popular since it can achieve high degrees of accuracy and it is very simple to train and tune. Although it has been applicable to a diverse range of problems in different areas, it has not found the same popularity for this particular problem.

The algorithm proposed by (Breiman, 2001) is trained using a bagging method which consists of sampling different subsets of the training data at random, fitting the model to these samples by using a random number of the features and aggregating the predictions using majority vote. The algorithm is able to perform a randomisation approach and introduce diversity using different features and by implementing the bagging method. This technique gives the algorithm the advantage of avoiding overfitting while handling a large number of input features.

Support Vector Machines (SVM)

Support Vector Machines are supervised learning algorithms able to produce nonlinear boundaries by building a linear boundary in the feature space transformed hyperplane. Its goal is to find the most optimal function that maximizes the distance between the nearest positive instances and the nearest negative instances. Figure 3.10 is a representation of this strategy:

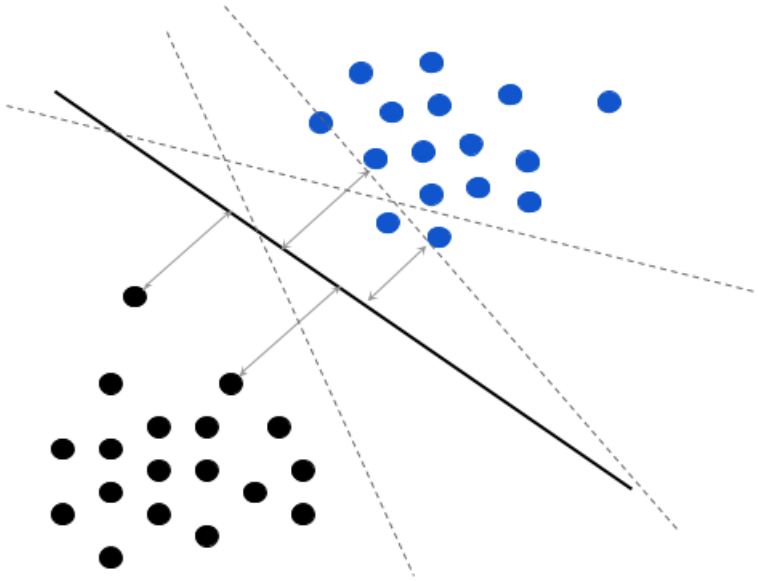

Figure 3.10: SVM: optimal separating hyperplane

Although this algorithm is relatively more complex than tree based algorithms, one of its advantages resides on the fact that the number of dimensions (features) does not affect its ability to find an optimal solution. However, it requires a significant amount of training time and memory.

Artificial Neural Networks (ANN)

Recently, Artificial Neural Networks have been a hot topic among researchers in machine learning and deep learning areas. This is a powerful algorithm able to outperform classic techniques such as Support Vector Machines and Random Forest.

In this study, a Multilayer Perceptron Neural Network was built. Usually, the architecture of this type of neural network consists of three layers where each of them are composed of units called neurons. Figure 3.11 illustrates the network structure implemented in this study:

- Input layer: contains a vector with the input features
- Hidden layer: contains the neurons that are interconnected to the input and output layers by weighted links

- Output layer: represents the model response

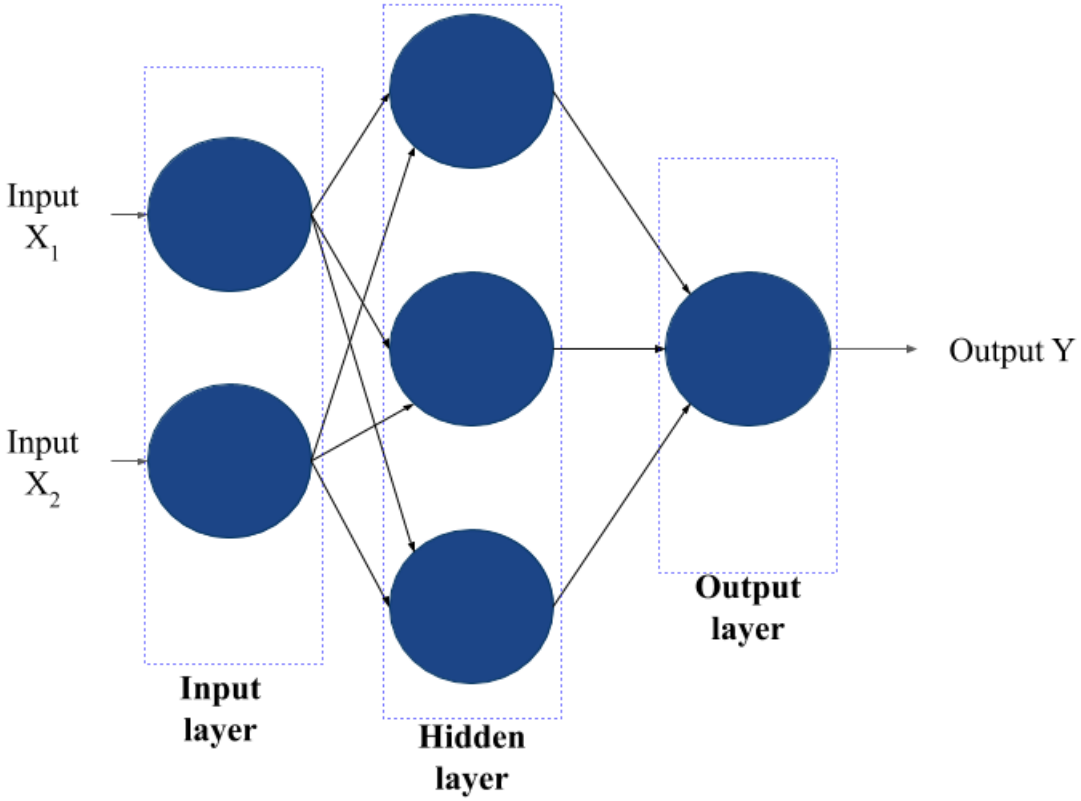

Figure 3.11: Multilayer Perceptron Neural Network

One advantage of these nonlinear statistical models is their capability to handle multiple responses smoothly. This makes this algorithm particularly well suited to process high dimensional data such as images. However, artificial neural networks are difficult to tune and they demand large amounts of computational processing power and memory usage.

Gradient Boosting (GB)

Gradient boosting is a type of supervised learning algorithm usually based on ensembled Decision Trees, in the same way as Random Forests. The idea behind this technique is that the model can be trained in a gradual and sequential fashion. The algorithm used gradient descent to optimize the loss function, meaning that the results

are adjusted by the gradient of the error between actual target and prediction. At each iteration, the model takes a step in the direction that minimizes the loss function.

Among the many advantages to use Gradient Boosting, resides in the fact that this algorithm allows the user to optimize a customized loss function.

3.3.5 Model Evaluation

This study has as a purpose the investigation of supervised learning algorithms that would allow the correct identification of HFO on seizure onset zones, as well as identifying those variables that contribute the most to this identification.

The determination of the goodness of fit of a classifier can be calculated by using the model trained on the training dataset to predict the classes on the test dataset and comparing the predictions to the true class values of the test dataset. There are several performance metrics that can make this comparison such as accuracy, F1-score, precision, recall, negative predictive value (NPV), sensitivity, specificity, receiver operating characteristics (ROC) curve analysis and area under ROC curve (AUC).

The metrics used to evaluate the model performance are crucial to assess which machine learning algorithm presents the greater predictive power. In line with previous studies in the literature, five criteria for evaluating the model performance were selected (Kubat, 2017). Table 3.2 presents a summary of the these metrics:

Metric	**Definition**
AUC	Measures the overall performance of a binary classifier
F1-Score	Measures the effectiveness of the model in terms of Precision and Recall
Accuracy	Measures the fraction of predictions that the model correctly identified
NPV	Measures the probability that predicted negative cases were actually negative
Recall	Measures the fraction of actual positive cases that were correctly identified as such
Precision	Measures the fraction of actual negative cases that were correctly identified as such

Table 3.2: Comparative table of the model performance metrics

While accuracy is presented in most research as one of the main metrics to evaluate model performance, it can be inaccurate to use only this metric when in presence of imbalance data as it is the case in question. When a particular class represents the majority of cases, an algorithm could achieve close to 100% of accuracy by predicting every instance as the majority class and it would be unsuccessful at identifying the minority class which is the main objective of the study. Hence, NPV, Recall, F1 Score and AUC were selected as additional metrics that can provide a good indication of the model performance when dealing with imbalanced data and are also useful to model comparison. This can be better observed when their formula is analysed:

1. **NPV:** It represents the probability that the classifier is right when labeling an instance as negative:

$$NPV = \frac{TN}{TN + FN} \tag{3.20}$$

2. **Precision:** It represents the probability that the classifier is right when labeling an instance as positive:

$$Precision = \frac{TP}{TP + FP} \tag{3.21}$$

3. **Recall:** it is the probability that the classifier would correctly predict a positive instance. Its formula is defined by:

$$Recall = \frac{TP}{TP + FN} \tag{3.22}$$

4. **Accuracy:** It is defined as the percentage of correctly classified instances and calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.23}$$

   Where:

   - TP = True positives cases
   - TN = True negatives cases
   - FP = False positives cases
   - FN = False negatives cases

5. F_1 **Score:** It is a criterion representing the harmonic mean of Precision and Recall, defined by the following formula:

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.24}$$

6. **AUC:** It stands for Area Under the Curve (ROC curve) and it measures the area underneath the ROC curve. It can be represented by a single scalar value that measures the overall performance of a binary classifier (Melo, 2013) and can take the values in the range [0, 1] in $\mathbb{R}$.

To define the algorithm with the best performance four metrics were selected: NPV, recall, ROC and F1 score. To select the model with the best performance

across all metrics, the absolute difference between a model's performance and the best performance for each metric was calculated. The model with the lowest relative difference was selected as the champion. The accuracy was calculated for each model for informational purposes since it is the most reported measure in the literature.

3.4 Evaluation of Designed Solution

As mentioned in the previous sections, model performance metrics such as F1 Score, Accuracy and AUC are good tools to compare models and assess which is the one that best performed the task that it was trained for. However, this study also requires a comparison between the Baseline model and Challenger was performed. The reason behind comparing different classification models is to understand whether any observed difference is statistically significant and not attributed to noise in the dataset.

Choosing the appropriate statistical test depends on the task the algorithm is trying to learn and the data distribution. The most common statistical test used in the literature to assess the difference of proportions is the independent t-test, however this is not the appropriate tool to assess the difference between the accuracy of two models. Instead, the McNemar test is proposed to determine if the difference observed is significantly different from the expected (Dietterich, 1998). The McNemar test is a non-parametric statistical test designed for paired nominal samples that can be applied to compare the performance metrics of two machine learning classifiers.

3.4.1 Hypothesis Test

Here the hypothesis test to assess the statistical significance of accuracy for the best performing algorithms used is presented. This hypothesis determines whether the difference in performance between the Baseline model and the proposed Challenger model are significant.

H0: There is no statistical significance between the models performance, i.e.: the probabilities p(A) and p(B) are the same.

H1:There is a statistical significance between the models performance.

Two significance levels were used to carry out the hypothesis test, 95% and 99%, therefore each hypothesis was rejected if p-value is less than α level (5% and 1%) respectively.

3.5 Strengths and Limitations of the proposed design

The experiment was designed having in mind that it should be easy to extend to other researches and that the proposed champion model could be applied to either EEG or iEEG data. The most relevant mainstream algorithms were selected due to their easy interpretability and understanding. Additionally, the fact that NPV is used as an evaluation metric which is important in many medical applications but not always reported in previous studies.

Another strength of the proposed solution is the fact that attributes extracted from different domains within the iEEG signal were used to train the model to identify pathological high frequency oscillations. Whereas prior research tends not to go in depth into feature creation, mostly focus on the machine learning algorithms.

Some of the limitations of the design include the fact that it still relies on iEEG data which is an invasive procedure that patients have to undergo. A possible solution would be to assess whether the application of the results obtained to EEG data is effective. In addition, since noise detection was not the core of this project there is an opportunity to increase the predictive power of the model by integrating it with more sophisticated solutions that are able to reduce the presence of artifacts in the data.

One of the potential issues of the design of this experiment is the limited amount of data. The predictive models are planned to be evaluated with information of only 8 patients. Therefore, if the patient selection had intrinsic bias the results of the evaluation could be less relevant in terms of the model ability to generalize to unseen data.

Chapter 4

Results, evaluation and discussion

A description of the experiment's implementation and the results obtained from this process, as well as their evaluation are discussed in this chapter. Having collected and processed the iEEG data, Baseline and Challenger models were trained to discriminate HFOs within the seizure onset zone using different supervised learning algorithms. The process consisted in defining those features that would be included in the Baseline and those included in the Challenger model, and then performing hyperparameter optimization to select the best settings for each algorithm. The full list of variables, the settings of the champion model (the one with the best performance) as well as the results obtained are presented in this section.

4.1 Baseline Model

The Baseline model was trained taking into consideration the most basic time domain variables used in the literature (Chen et al., 2017; Dian et al., 2015). Table 4.1 presents the selected features:

Basic Time Domain Features	
Min	Max
Mean	Std
Skewness	Kurtosis
Coef. Variation	Entropy

Table 4.1: Baseline models - input features

For each algorithm, an iterative approach was chosen to carry out the hyperparameter optimization during the training and validation phase. This optimization consists of choosing a set of optimal hyperparameters that define the model architecture and searching for the ideal combination of such hyperparameters. In this work, the Grid search method was used to identify the best set of hyperparameters by training a model for each combination of all the parameter values specified for each algorithm. Then the best set of parameters was selected based on the model that produced the best results employing a 10-fold cross validation.

Table 4.2 details the different hyperparameters that were optimized during this iterative process which is described by the following steps:

1. Iteration 1: model is trained with the default hyperparameters and the Baseline features

2. Iteration 2: manually selection of different values for the several algorithm's hyperparameters to understand which are the few important parameters

3. Iteration 3: perform a grid search with a 10-fold cross validation over the most important parameters and select the optimal settings

Algorithm	**Hyperparameters**
Decision Tree	criterion min samples leaf min samples split max features
Random Forest	n estimators criterion min samples leaf max features
Gradient Boosting	learning rate n estimators min samples leaf max features
Support Vector Machine	gamma C
Artificial Neural Networks	hidden layer sizes activation alpha batch size max iter

Table 4.2: Hyperparameters tested for each algorithm

Given the lack of consensus in the literature about the best algorithm to train and solve for this type of problem, four different algorithms were selected based on previous research and the model with the best performance was selected as the champion model. Table 4.3 presents the performance metrics obtained by these algorithms:

Model	**Precision**	**Recall**	**AUC**	**Accuracy**	**F1**
DT (Baseline)	47.5%	22.5%	60.1%	91.4%	0.31
RF (Baseline)	64.8%	20.2%	59.6%	92.4%	0.32
GB (Baseline)	68.9%	10.0%	54.8%	92.1%	0.17
ANN (Baseline)	71.7%	11.2%	55.4%	92.2%	0.19
SVM (Baseline)	**19.4%**	**62.7%**	**69.5%**	**75.2%**	**0.30**

Table 4.3: Baseline model. Performance metrics: Decision Tree (DT); Random Forest (RF); Gradient Boosting (GB); Support Vector Machine (SVM); Artificial Neural Network (ANN)

As it can be seen from the results obtained, the Random Forest algorithm presents the highest F1 Score (0.32) and accuracy (92.4%). As mentioned in the previous section, when a dataset is imbalanced looking at accuracy is not recommended and instead the metrics such as Precision, Recall and NPV should be considered to select the best model.

In the case of the Artificial Neural Network, it presents a high value of precision but low recall, meaning that while the model is good at predicting positive cases, it only identifies a small number of the total positives. In this context, when the model identifies a peak as pathological HFO (pHFO) it is most likely to be a real pHFO but not all of the existing pHFO are being caught by the model.

On the other hand, the highest recall is achieved by the SVM model despite lower values of precision. This means that the model is predicting most positive cases correctly but it is also predicting some of the negative cases as positive. This model also presents the highest AUC metric overall. Figure 4.1 presents the ROC of each model.

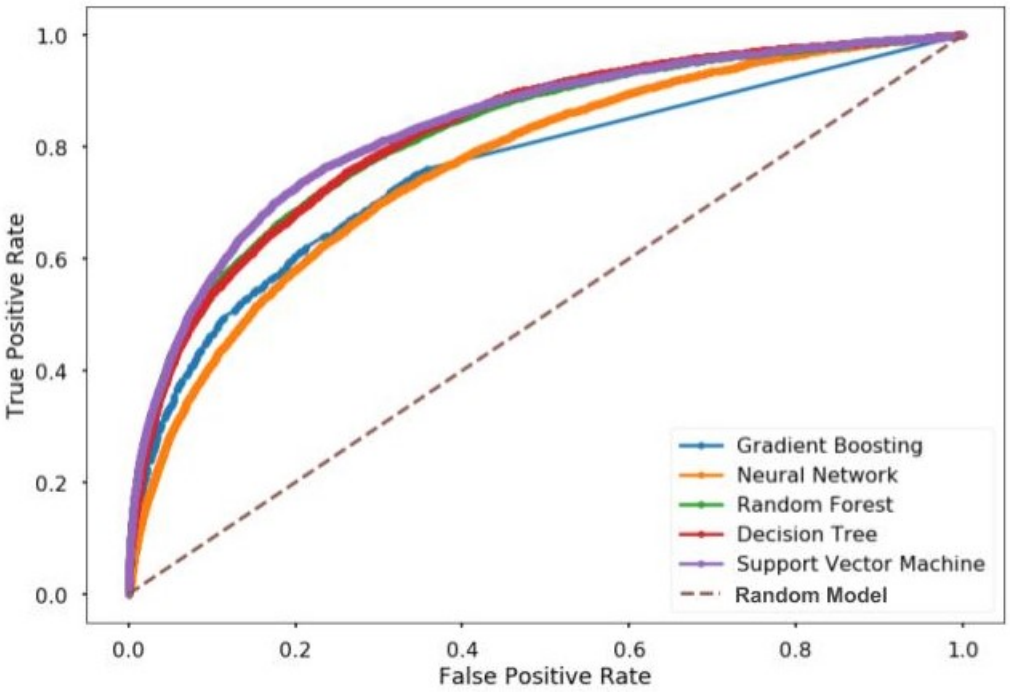

Figure 4.1: ROC curves of baseline models

Although high values of accuracy and AUC are important in every machine learning model, in most machine learning medical applications higher values of recall are preferred than high values of precision. This is due to the fact that predicting some negative cases as positive would not present major consequences, these cases will be reviewed by doctors and then classified correctly as negative.

Another important metric is the NPV, which is a measure of the clinical relevance of the results. The NPV uses the prevalence of a condition to determine the probability that the predicted negative cases are in fact negative. This metric is important because if a pHFO is predicted to be non-pathological (nHFO) it would have a major negative impact (doctors will not take this for further review). Table 4.4 presents the NPV values achieved by each model:

Model	**NPV**
DT (Baseline)	93.3%
RF (Baseline)	93.2%
GB (Baseline)	92.4%
ANN (Baseline)	92.5%
SVM (Baseline)	**95.7%**

Table 4.4: Negative Predictive Values of baselines models

As mentioned in Section 3.3.5, the champion model was selected taking into consideration Recall, AUC, F1 Score and NPV metrics. The model with the best performance across these metrics was a Support Vector Machine (SVM) with the following settings:

Hyperparameters	**Values**
Kernel	Radial Basis Function (RBF)
gamma	1
C	10

Table 4.5: Baseline models. Best hyperparameters settings

Next, an analysis of the features' importance was completed. However, not all algorithms can provide a measure of how important a feature is, this being the case of Support Vector Machine and Artificial Neural Networks. Therefore, the importance of the features in the three based algorithms was considered as a proxy to identify the most relevant. Table 4.6 presents the feature for each model by order of importance:

DT	**RF**	**GB**
std	std	std
skewness	CV	kurtosis
CV	skewness	CV
kurtosis	kurtosis	skewness
	max	max

Table 4.6: Baseline models. Features selected by decision tree (DT), random forest (RF) and gradient boosting (GB) by order of importance.

After analysing the results, it becomes clear that standard deviation (std) within the signal is the most relevant feature when it comes to identify pathological HFOs. Also, the skewness and the coefficient of correlation are very important to the correct prediction.

4.2 Challenger Model

To test whether time domain and time-frequency domain features are better predictors of pHFO a set of Challenger models were built. This second set of models were trained using a combination of traditional features, time domain and time-frequency domain attributes. The complete list of features tested is presented in Table 4.7:

Complete set of features		
Min	Max	Mean
Std	Skewness	Kurtosis
Coef. Variation	Entropy	Energy
Peak Maximum Avg	Peak to Notch	Amplitude Envelope
KL Dist	Entropy	Power Band Ratio
Number of Peaks		

Table 4.7: Input features of challenger models

In addition, the same set of algorithms used to train the set of Baseline models were also implemented to train the Challenger models. The set of optimal hyperparameters was obtained following a similar strategy to the Baseline set. Initially, the best set of parameters from the Baseline models were implemented as a first iteration to train the Challenger model. Then, based on this initial set of hyperparameters a Grid search implementing a 10-fold cross validation technique was used to test different values and therefore obtain the best hypeparameters for each algorithm. Finally, five Challenger models were trained and validated using the optimized values. The performance metrics of each model is presented in table below:

Model	**Precision**	**Recall**	**AUC**	**Accuracy**	**F1**
DT (Challenger)	64.3%	42.5%	70.2%	93.2%	0.51
RF (Challenger)	81.9%	41.5%	70.3%	94.3%	0.55
GB (Challenger)	79.6%	35.8%	67.5%	93.9%	0.49
ANN (Challenger)	80.7%	36.3%	67.7%	93.9%	0.50
SVM (Challenger)	**33.5%**	**77.0%**	**81.5%**	**85.3%**	**0.47**

Table 4.8: Challenger model. Performance metrics: Decision Tree (DT); Random Forest (RF); Gradient Boosting (GB); Support Vector Machine (SVM); Artificial Neural Network (ANN)

The Random Forest algorithm presents the highest F1 Score (0.55), accuracy (94.3%) and precision (81.9%) across all models, similar to what was observed in the Baseline model setup. These results prove how powerful the Random Forest algorithm is and why it is one of the most popular choices among many researchers.

Nonetheless, when recall and AUC values are analysed, the best model that arises is a SVM model which achieved an AUC of 81.5% and recall of 77%. Both metrics indicate the model is very good predicting correctly the positive cases in the population. The ROC curve for each model is displayed in Table 4.2.

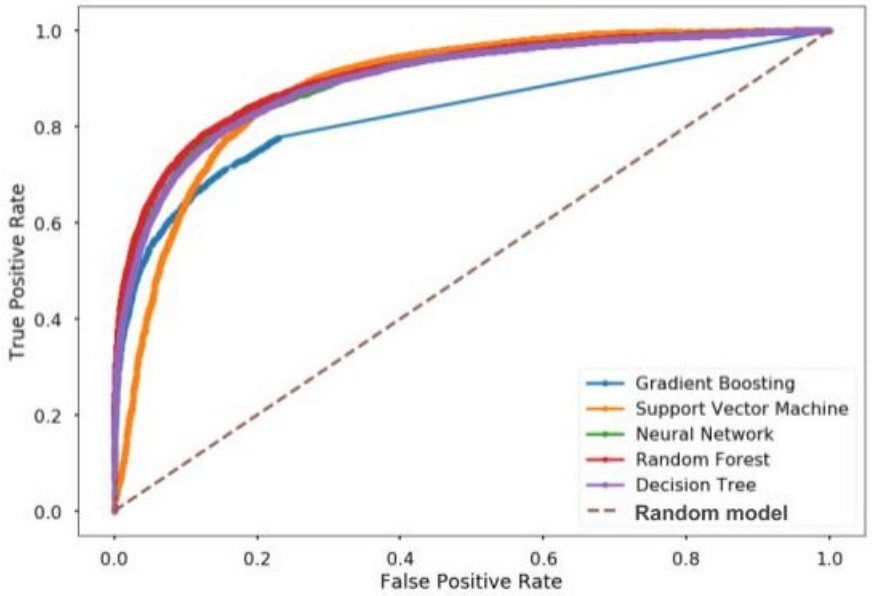

Figure 4.2: ROC curves of challenger models

As mentioned earlier, NPV is also a highly important metric that needs to be reported when dealing with this type of problem. Based on the results achieved by each algorithm presented in Table 4.9, it can be observed that SVM model has the highest predictive power when it comes to identifying the negative cases.

Model	**NPV**
DT (Challenger)	94.9%
RF (Challenger)	95%
GB (Challenger)	94.4%
ANN (Challenger)	94.5%
SVM (Challenger)	**97.6%**

Table 4.9: Negative Predictive Values of challenger models

Based on the results obtained, the model with the best performance across the different metrics was a Support Vector Machine with the following settings:

Hyperparameters	**Values**
Kernel	Radial Basis Function (RBF)
gamma	50
C	100

Table 4.10: Final hyperparameters settings of SVM model.

When the features' importance is examined in each model, it shows that the frequency and time-frequency variables (e.g.: energy, entropy, KL distance, number of peaks, etc.) have more relevance in the models than the original basic time domain features. Furthermore, attributes such as standard deviation and coefficient of variation also remain important over the models. Table 4.11 details the list of features by importance for each model:

DT	**RF**	**GB**
std	CV	num peak
entropy	std	entropy
CV	entropy	CV
KL dist	energy	energy
kurtosis	KL dist	std
energy	peak to notch	kurtosis
num peak	num peak	KL dist
PBR	kurtosis	PBR
skewness	skewness	peak to notch
peak to notch	PBR	skewness
max	amplitude envelope	amplitude envelope
	max	
	max peak avg	

Table 4.11: Challenger models. Features selected by decision tree (DT), random forest (RF) and gradient boosting (GB) by order of importance

4.3 Model comparison

One champion model was selected from the set of Baseline models and another champion from the set of Challenger models. The following table summarizes the performance metrics for both champions selected and Figure 4.3 displays the ROC curve for each model:

Model	**Precision**	**Recall**	**AUC**	**Accuracy**	**F1**
Best Baseline (SVM)	19.4%	62.7%	69.5%	75.2%	0.30
Best Challenger (SVM)	33.5%	77.0%	81.5%	85.3%	0.47

Table 4.12: Comparison of best performing baseline model (applying time-frequency domain features) and best challenger model (applying all sets of features listed in Table 3.1).

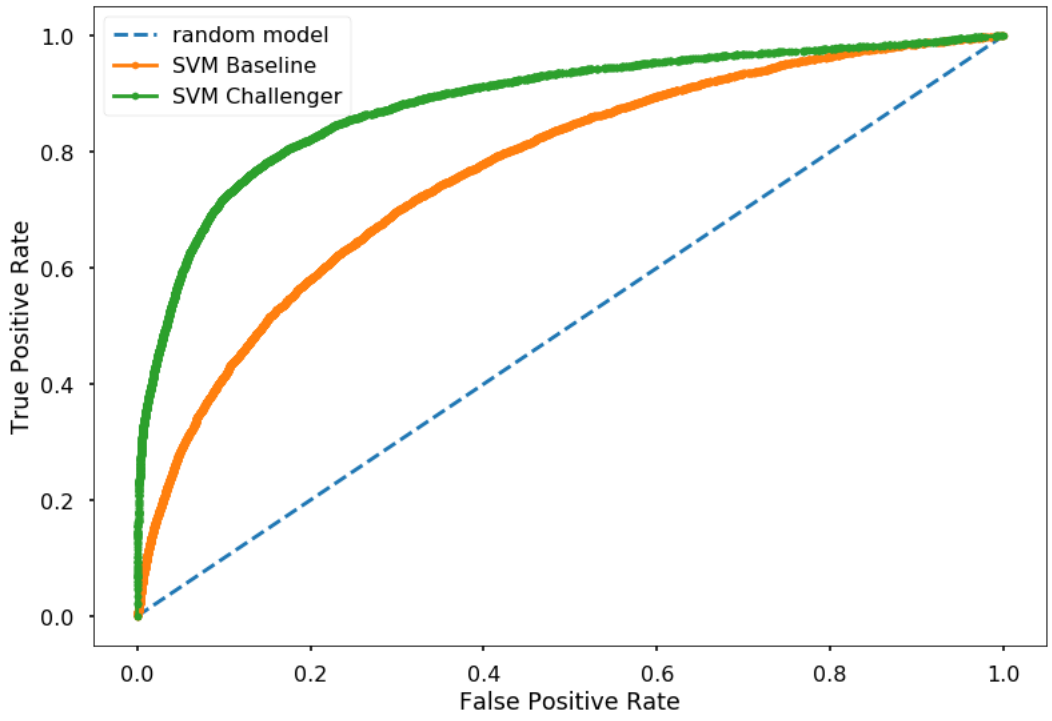

Figure 4.3: ROC Curve of SVM challenger and SVM baseline models.

As it can be observed, the performance of the champion Challenger model is quite superior in most metrics. The F1 Score of the Challenger is 57% higher than the Baseline, as well as its precision is 72% higher. In addition, the AUC of the cham-

pion Challenger model is only a 17% better than the champion Baseline. Overall, it suggests that introducing the frequency and time-frequency attributes improved the performance of the algorithm in identifying the pHFOs.

As explained earlier, SVM algorithm does not provide guidance with regards to the features that contribute the most to the model. Therefore, the output from the other algorithms (decision tree, random forest and gradient boosting) were used to identify the best predictors by taking into consideration the ranking of each attribute when ordered by its importance. Table 4.13 presents a summary of the top 5 features in each model:

Baseline	**Challenger**
std	Entropy
CV	CV
Skewness	std
Kurtosis	Energy
max	KL dist

Table 4.13: Top 5 best performing features in each model based on Decision tree, Random Forest and Gradient Boosting results.

Since the decision tree, random forest and gradient boosting are tree based algorithms, the most important features can be selected computing the relative rank of an attribute when used as a decision node in the tree. Usually, attributes used at the top of the tree tend to contribute to the final prediction of the majority of the input samples (Breiman, 2001). The five most important features of the baseline and challenger model were selected combining their raking in each of the three algorithms mentioned above. Tables 4.6 and 4.11 presented the ranking of the attributes for each algorithm in the baseline and challenger model respectively.

Also, the density distribution of the top five best performing features were analysed (Figure 4.4 - Figure 4.8). From this analysis it can be observed that the features capturing the oscillations in the SOZ demonstrate a long tail, especially for Entropy and

Energy features which is compatible to what has been seen in the literature (Cimbalnik et al., 2019).

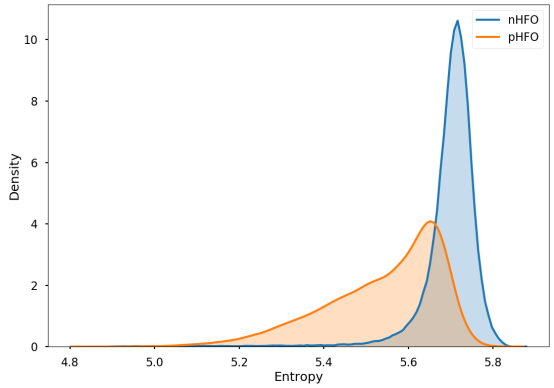

Figure 4.4: Entropy distribution

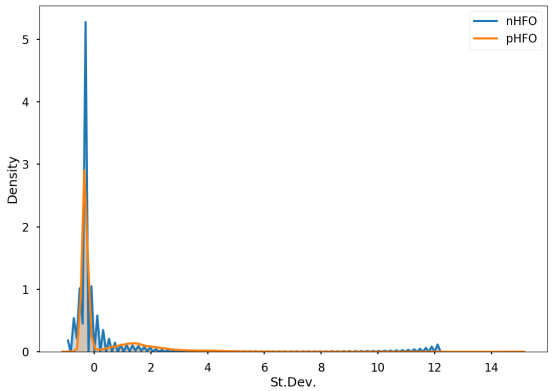

Figure 4.5: St. Deviation distribution

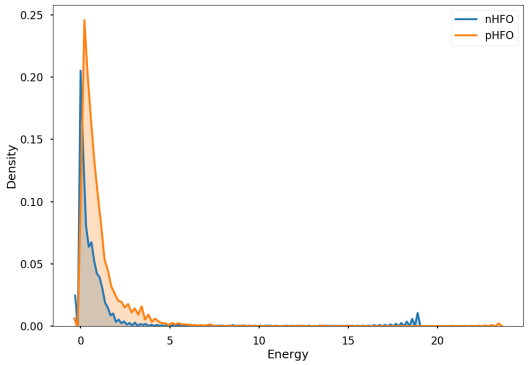

Figure 4.6: Energy distribution

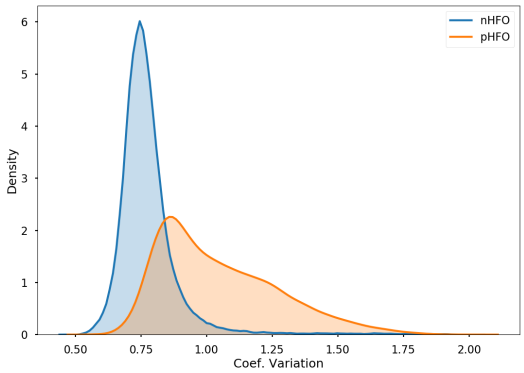

Figure 4.7: CV distribution

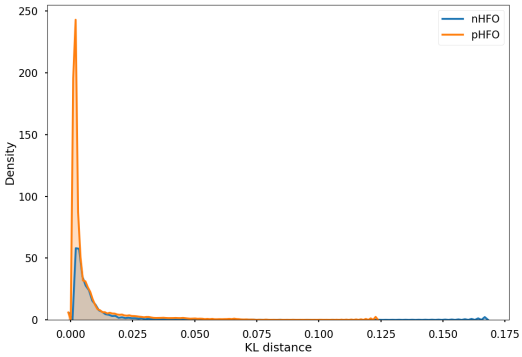

Figure 4.8: KL distance distribution

Additionally, when compared to results achieved by previous researchers, it is observed that the proposed champion model accomplished excellent values of NPV and good values of AUC. However, the F1 score is not as high as expected due to low precision of the model. Table 4.14 presents the comparison against similar models presented in the literature:

	Methodology		**Best Performance**		
Author	**Features**	**Classifier**	**AUC**	**NPV**	**F1 Score**
Akter et al. (2020)	features based on entropy	SVM	83%	N/A	0.46
Chen et al. (2017)	Max, Min, Mean, Std, Skewness, Kurtosis, Energy	SVM	92%	91.8%	0.92
Cimbalnik et al. (2019)	HFO, Univariate and Bivariate features	SVM	N/A	80%	0.89
Proposed Model	Entropy, CV, Std, Energy, KL dist	SVM	77%	97.6%	0.47

Table 4.14: Comparison of proposed solution to results achieved by previous research

The results obtained suggest that the precision of the algorithm could be improved with no detriment of the recall metrics, and thus the F1 score achieved should increase significantly.

4.3.1 McNemar test

A paired sample McNemar statistical test was conducted to compare whether the difference in the performance obtained by the two different models are actually statistically significant. The results of the test suggest that the difference between these models was significant, X^2 (2, N = 52842) = 551, $p < .001$. The predictive power of the Challenger model is higher than the predictive power achieved by the Baseline

model, suggesting that the features added are indeed better predictors. Therefore, the statistically significant result indicates there is evidence to support rejecting the null hypothesis presented in Section 3.1.

4.4 Evaluation

To evaluate the model capability to generalize to unseen data, the champion model was implemented on 8 patients that did not participate in the training, validation nor test phases. Table 4.15 presents the performance metrics of the model for each patient:

Patient	**Precision**	**Recall**	**AUC**	**Accuracy**	**F1**
Patient 1	33.9%	68%	77.8%	85.9%	0.45
Patient 2	24%	46%	70%	92.2%	0.32
Patient 3	1.6%	59%	54.4%	50%	0.03
Patient 4	10.1%	15.5%	53.7%	87.8%	0.12
Patient 5	4.2%	44.7%	59.1%	72.7%	0.08
Patient 6	10.6%	69.1%	61.9%	55.8%	0.18
Patient 7	5.9%	04.5%	51.3%	95.6%	0.05
Patient 8	0%	0%	49.6%	98%	0

Table 4.15: Performance metrics for extra 8 individual patients of SVM challenger model.

The results achieved by the model are very heterogeneous, they suggest that the model did a good work predicting pHFO on half of the patients analyzed. Although, the model accomplished a very good AUC for patients 1,2 and 6, its performance for patients 7 and 8 is not different than a random guess.

As can be observed in Table 4.16, the model performance in terms of negative predictive values is relatively good even for those patients where the overall performance was poor.

Patient	**NPV**
Patient 1	96.7%
Patient 2	97.7%
Patient 3	98.8%
Patient 4	94.9%
Patient 5	98.1%
Patient 6	95.8%
Patient 7	97.4%
Patient 8	98.7%

Table 4.16: Negative predictive values for extra 8 individual patients of SVM challenger model.

A possible explanation for the acceptable values of NPV could be due to the fact that the algorithm overestimates the number of HFO in the seizure onset zone, as can been seen in Figure 4.9 that shows the number of predicted pHFO (orange bar) and the total number of pHFO identified visually by doctors (blue bar).

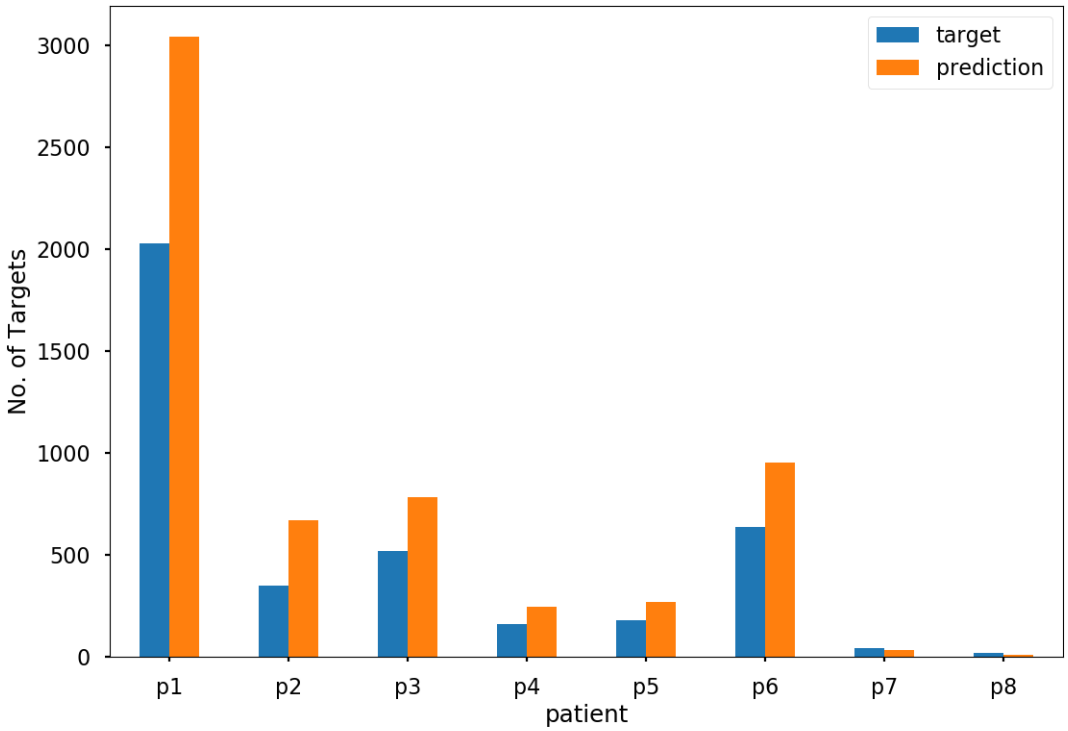

Figure 4.9: Patient-specific target detection. Each bar indicates the number of targets

Nonetheless, when the number of correct predictions per patient is analysed (Figure 4.10) it is observed that the model has missed the true pHFO in 4 out of the 8 patients analysed which it is in accordance with the low precision of the model.

For Patient 1 and 6 the algorithm over detected 68% of the total targets, however it also predicted 2 and 6 times more events than the actual numbers. On the other hand, the model presented its worst performance on patients 7 and 8 where it identified less than 5% of the total events.

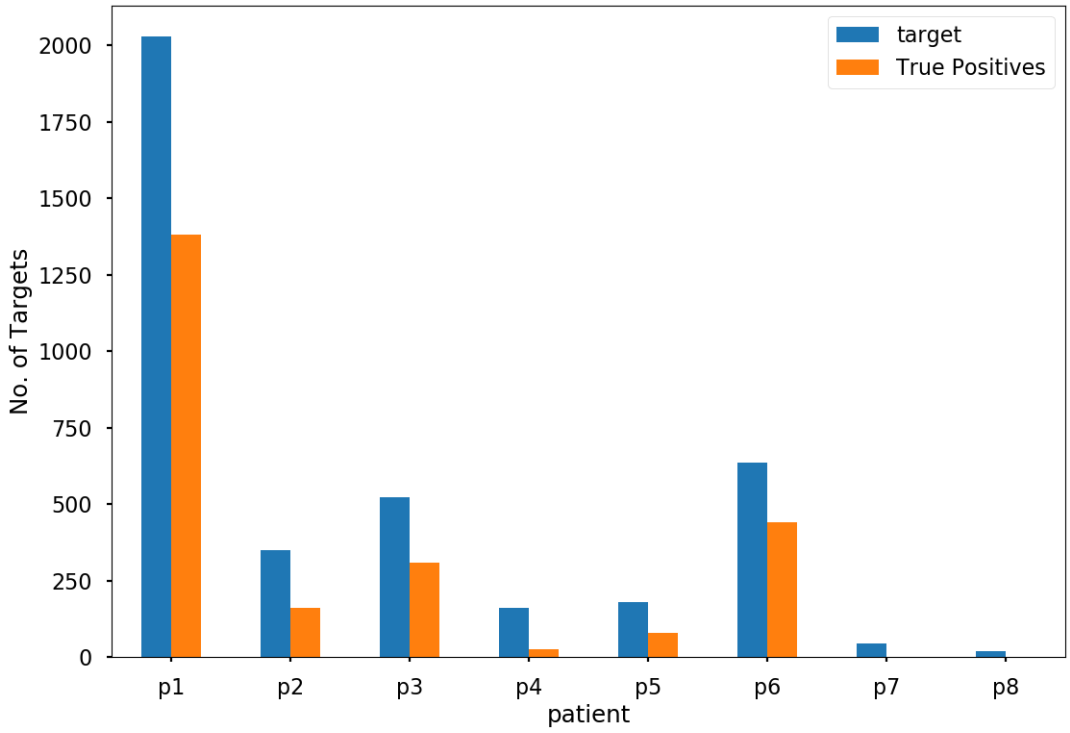

Figure 4.10: Patient-specific target detection. Each bar indicates the number of targets (blue corresponds to actual target and orange to those correctly predicted)

There are additional factors that might be relevant to this analysis but were not taken into consideration, for example the number of channels used to capture the iEEG. It was observed that in the cases where less channels were used, the model presented an inferior performance, as can be seen in Figure 4.11. To confirm this hypothesis, data of the thirty three patients was used and a Pearson's correlation test was conducted to assess the relationship between the model AUC performance and the number of channels. The results suggested that there is a strong statistically significant correlation between the two variables (r =.7417, n = 33, p=.0351). However, it must be considered that the sample size is quite small.

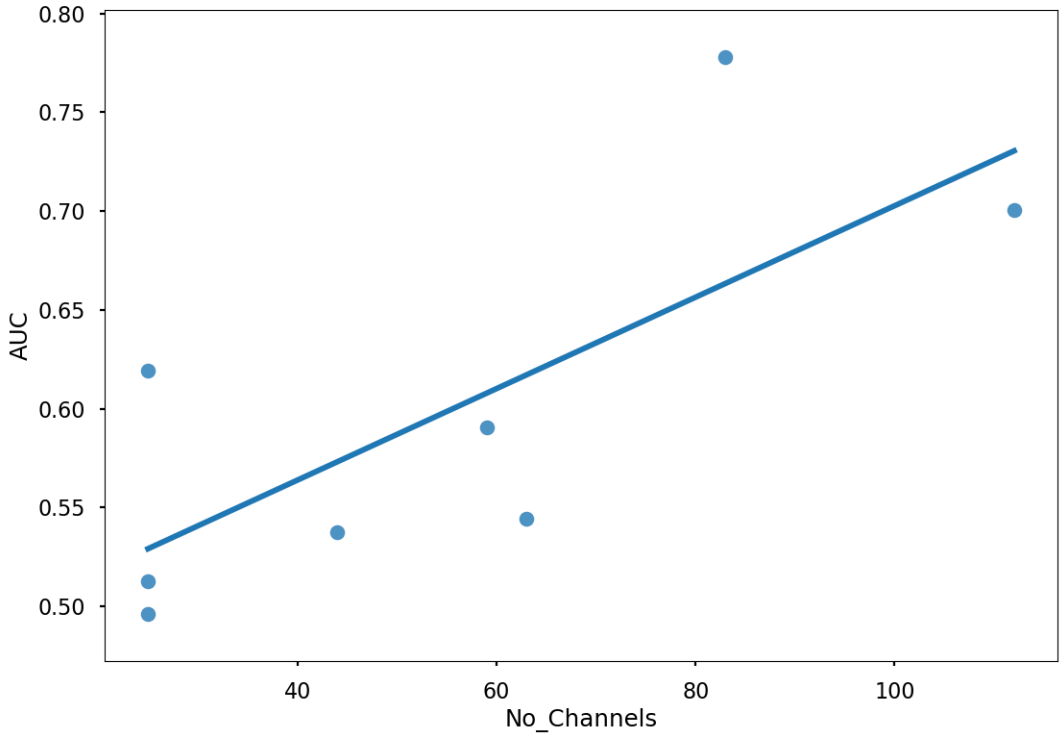

Figure 4.11: Relationship between number of channels used to record iEEG and AUC performance of the model)

4.5 Discussion

Some of the limitations and strengths of the proposed champion model are discussed in this section.

With regards to the strengths, the fact that the model was trained with a larger dataset than what has been seen in the literature, suggests that data points based on signals with different shapes were included and therefore the model potential to generalize should be better. Also, the negative predicted value was reasonably better than those achieved by previous studies.

An additional strength of the model is that the features used to improve the model performance did not require additional data collection. All the features presented were extracted from the same iEEG recordings.

Although the results obtained are promising, there are a few limitations that are

worth mentioning. First, the running time of the SVM algorithm was significantly higher than for other algorithms which requires an analysis of the trade off between better performance and training time needed. Second, the analysis of adding these features individually was not included in this research, as well as different combinations of multiple subsets of the attributes.

Finally, the model was not trained to detect seizures happening in real time, it was only trained to identify HFO located in the seizure onset zone. However, translating this model into a real time seizure detection application should not be difficult based on the assumption that brain waves would have similar shape

Chapter 5

Conclusion

5.1 Research Overview

Epilepsy is the second prevalence neurological disease in the world, affecting millions of people worldwide. In most cases this condition is treated with medication, yet drug treatment is not successful for many patients whose need to go under surgery to control the seizures produced by this disorder.

This research attempts to expand the knowledge about this subject by combining different features and classifiers selected from the literature. In this study, features from different domains were extracted from intracranial electroencephalography (iEEG) signals are studied in order to determine their fitness to identify pathological high frequency oscillations (HFO) by training a series of supervised machine learning algorithms to perform this task. The discrimination between pathological (pHFO) and physiological HFO would help epileptologists to focus their resources on the seizure onset zone (SOZ) in the brain.

5.2 Problem Definition

The current gold standard to identify the seizure onset zone (SOZ) in the brain is based on visual inspection of the iEEG captured through a surgical procedure. Therefore, most researchers have analysed HFO as a potential biomarker. However pathological

and physiological oscillations can occur across different brain areas and in order to successfully identify the pathological HFO (pHFO) on the SOZ it is important to understand which variables successfully represent the pHFO characteristics.

In addition, many different machine learning approaches have been used to tackle this problem. Some researchers have adopted unsupervised learning approaches using mostly clustering techniques based on Gaussian Mixed Models. On the other hand Support Vector Machines have been very popular among those studies using a supervised learning approach.

All the reasons presented above explain why this is not a trivial task and requires a deep understanding of the different methods that can be used to solve this problem. Thus, the main goal of this research was to identify the most significant iEEG signal features for the detection of pathological HFO on the SOZ. Secondly, a statistical model using supervised machine learning algorithms was developed to detect such pHFO in epilepsy patients.

5.3 Design, Evaluation & Results

The presented analysis was motivated by the need to identify those features extracted from iEEG signal data that are statistically significant to discriminate HFO in the seizures onset zones. The hypothesis behind this is that supervised learning models trained using less traditional attributes will show better performance than the models that use only traditional attributes. Where less traditional features are frequency domain attributes (e.g.: entropy, energy, etc.) and time-frequency domain attributes (e.g. number of peaks, power spectral density, Kullback-Leibler distance, etc.). Usually, the traditional features refer to basic statistical time domain variables such as mean, skewness, standard deviation, etc.

As a result, a total of five Baseline models were trained using traditional features and employing five different supervised learning algorithms:

1. Decision Tree

2. Random Forest

3. Gradient Boosting

4. Support Vector Machine

5. Artificial Neural Network

Likewise, five different Challenger models were built taking into consideration both traditional and non-traditional attributes and making use of the same set of algorithms.

Data of 25 out of the 33 patients were used to train, validate and test the machine learning models. Also, a grid search with a 10-fold cross validation approach was employed to optimize the hyperparameters of each algorithm. Also, to assess the champion model's ability to generalize to unseen data, the model was applied to the remaining 8 patients.

From the set of Baseline models, a SVM (kernel = rbf, C= 10, gamma = 1) was the best performing model achieving 75.2% accuracy, 69.5% AUC and 95.7% NPV. However these values suggest that there is opportunity to improve the predictive power of the model. Furthermore, the best Challenger model was also a SVM (kernel = rbf, C= 100, gamma = 50) algorithm. This model presented a 23% higher recall (77%), 17% higher AUC (82%), 13% higher accuracy(85%) and the negative predictive value rate also improved to 97.6%.

With respect to the features with the highest importance, these could not be determined from the SVM algorithm since it projects the features hyperplane to an infinite dimension. However, the outputs from the Random Forest, Gradient Boosting and Decision Tree were taken into consideration to define which features were the better predictors. From this analysis, it was observed that entropy, coefficient of variation, standard deviation, energy and Kullback-Leibler distance were the five attributes with the most importance to the models. This suggests that frequency and time-domain features contribute to the discrimination of pHFO in the seizure onset zone.

Finally, a McNemar test was conducted to assess whether the difference between the performance of the models is statistically significant. The results of the test suggest that the difference between these models was significant, X^2 (2, N = 52842) = 551,

$p < .001$ and that there is evidence to support rejecting the null hypothesis stated in Section 3.1.

5.4 Contributions and impact

The results from the research carried out suggest that the inclusion of frequency domain and time-frequency domain variables to train machine learning models to discriminate between pathological and physiological HFO improves the model performance. This is one of the few studies that have combined different attributes extracted from iEEG data collected to train a machine learning model able to identify pHFO in the seizure onset zone.

In addition, when it comes to assessing the best algorithm the SVM proved to be the best option in terms of getting high values of negative predictive value (NPV), AUC and Recall. The recommended metrics used to evaluate machine learning algorithms when applied to medical cases are recall and NPV. High values of recall indicate that the algorithm is successful in classifying most positive cases. Furthermore, high values of NPV imply that the algorithm is identifying negative cases correctly.

The fact the SVM algorithms performed better in both Baseline and Challenger problems, indicates that its ability to learn non-linear decision boundaries is really powerful when applied to iEEG signal data. This finding adds to the large body of knowledge in the field of machine learning applied to neurological sciences and supports why many researchers have used it in the past.

When compared to other studies, it was observed that the champion model presented similar levels of accuracy and a higher NPV, which is something to be outlined since NPV is a crucial metric when working with medical applications.

Moreover, besides the findings obtained by this research, it is also important to emphasize that this study used a larger dataset than what has been seen in the literature (presented in Table 2). The findings are based on 33 patients while most of previous research have used less than 20 patients.

Although the analysis presented in this study is limited to the use of HFO as

biomarkers, the approach proposed and the findings presented show value. They emphasize the importance of using machine learning techniques to solve important medical problems and help doctors on time consuming tasks.

5.5 Future Work & recommendations

The work presented aimed to fill one of the many gaps in the literature, hence there is an opportunity to continue with this analysis complementing it with additional data. For example, previous research suggests that features extracted from other biomarkers could represent a huge improvement regarding the identification of seizure onset zones (Cimbalnik 2019 y Varatharajah,). There are several examples of features that could be included such as phase locking value (PLV) and cross frequency coupling (CFC) based features.

Although multiple supervised machine learning algorithms have been tested in this study, future work could focus on the adoption of deep learning techniques to identify the relevant oscillations on the SOZ.

Finally, how to discriminate between pathological and physiological HFOs generated by the human brain is a task that has yet to be completely understood. Hopefully, this work could set the base to future developments into real-time iEEG data processing and SOZ discrimination, which could reduce the time required for monitoring and could lead to real-time seizure identification.

References

Akter, M. S., Islam, M. R., Iimura, Y., Sugano, H., Fukumori, K., Wang, D., ... Cichocki, A. (2020). Multiband entropy-based feature-extraction method for automatic identification of epileptic focus based on high-frequency components in interictal iEEG. *Scientific Reports*, *10*(1), 7044. doi: 10.1038/s41598-020-62967-z

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neuroscience*, *2014*, 1–7. doi: 10.1155/2014/730218

Alper, K., Raghavan, M., Isenhart, R., Howard, B., Doyle, W., John, R., & Prichep, L. (2008). Localizing epileptogenic regions in partial epilepsy using three-dimensional statistical parametric maps of background EEG source spectra. *NeuroImage*, *39*(3), 1257–1265. doi: 10.1016/j.neuroimage.2007.09.041

Amiri, M., Frauscher, B., & Gotman, J. (2016). Phase-amplitude coupling is elevated in deep sleep and in the onset zone of focal epileptic seizures. *Frontiers in Human Neuroscience*, *10*(August), 12. doi: 10.3389/fnhum.2016.00387

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi: https://doi.org/10.1023/A:1010933404324

Channel-Wise Characterization of High Frequency Oscillations for Automated Identification of the Seizure Onset Zone. (2020). *IEEE Access*, *8*, 45531–45543. doi: 10.1109/ACCESS.2020.2978290

Chen, D., Wan, S., Xiang, J., & Bao, F. S. (2017). A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG. *PLoS ONE*, *12*(3), 1–21. doi: 10.1371/journal.pone.0173138

Chua, K. C., Chandran, V., Acharya, R., & Lim, C. M. (2008). Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: A comparative study. *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology"*, 3824–3827. doi: 10.1109/iembs.2008.4650043

Cimbalnik, J., Klimes, P., Sladky, V., Nejedly, P., Jurak, P., Pail, M., ... Worrell, G. (2019). Multi-feature localization of epileptic foci from interictal, intracranial EEG. *Clinical Neurophysiology*, *130*(10), 1945–1953. doi: 10.1016/j.clinph.2019.07.024

Cimbalnik, J., Kucewicz, M. T., & Worrell, G. A. (2016). Interictal high-frequency oscillations in focal human epilepsy. *Current opinion in Neurology*, *29*(1), 175–181. doi: 10.14574/ojrnhc.v14i1.276.Using

De Tisi, J., Bell, G. S., Peacock, J. L., McEvoy, A. W., Harkness, W. F., Sander, J. W., & Duncan, J. S. (2011). The long-term outcome of adult epilepsy surgery, patterns of seizure remission, and relapse: A cohort study. *The Lancet*, *378*(9800), 1388–1395. doi: 10.1016/S0140-6736(11)60890-8

Dian, J. A., Colic, S., Chinvarun, Y., Carlen, P. L., & Bardakjian, B. L. (2015). Identification of brain regions of interest for epilepsy surgery planning using support vector machines. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2015-Novem*, 6590–6593. doi: 10.1109/EMBC.2015.7319903

Dietterich, T. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1924. doi: 10.1162/089976698300017197

Dümpelmann, M., Jacobs, J., & Schulze-Bonhage, A. (2015). Temporal and spatial characteristics of high frequency oscillations as a new biomarker in epilepsy. *Epilepsia*, *56*(2), 197–206. doi: 10.1111/epi.12844

Edakawa, K., Yanagisawa, T., Kishima, H., Fukuma, R., Oshino, S., Khoo, H. M., ... Yoshimine, T. (2016). Detection of Epileptic Seizures Using Phase-Amplitude Coupling in Intracranial Electroencephalography. *Scientific Reports*, *6*(May), 1–8. doi: 10.1038/srep25422

Elahian, B., Yeasin, M., Mudigoudar, B., Wheless, J. W., & Babajani-Feremi, A. (2017). Identifying seizure onset zone from electrocorticographic recordings: A machine learning approach based on phase locking value. *Seizure*, *51*, 35–42. doi: 10.1016/j.seizure.2017.07.010

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., ... Wiebe, S. (2014). ILAE Official Report: A practical clinical definition of epilepsy. *Epilepsia*, *55*(4), 475–482. doi: 10.1111/epi.12550

Frauscher, B., Bartolomei, F., Kobayashi, K., Cimbalnik, J., van 't Klooster, M. A., Rampp, S., ... Gotman, J. (2017). High-frequency oscillations: The state of clinical research. *Epilepsia*, *58*(8), 1316–1329. doi: 10.1111/epi.13829

Gabor, A. J., Leach, R. R., & Dowla, F. U. (1996). Automated seizure detection using a self-organizing neural network. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 257–266. doi: 10.1016/0013-4694(96)96001-0

Gaspard, N., Alkawadri, R., Farooque, P., Goncharova, I. I., & Zaveri, H. P. (2014). Automatic detection of prominent interictal spikes in intracranial EEG: Validation of an algorithm and relationsip to the seizure onset zone. *Clinical Neurophysiology*, *125*(6), 1095–1103. doi: 10.1016/j.clinph.2013.10.021

Ghosh-Dastidar, S., & Adeli, H. (2009). A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection. *Neural Networks*, *22*(10), 1419–1431. doi: 10.1016/j.neunet.2009.04.003

Gliske, S. V., Irwin, Z. T., Davis, K. A., Sahaya, K., Chestek, C., & Stacey, W. C. (2016). Universal automated high frequency oscillation detector for real-time, long term EEG. *Clinical Neurophysiology*, *127*(2), 1057–1066. doi: 10.1016/j.clinph.2015.07.016

Guirgis, M., Chinvarun, Y., Del Campo, M., Carlen, P. L., & Bardakjian, B. L. (2015). Defining regions of interest using cross-frequency coupling in extratemporal lobe epilepsy patients. *Journal of Neural Engineering*, *12*(2). doi: 10.1088/1741-2560/12/2/026011

Guo, L., Rivero, D., & Pazos, A. (2010). Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. *Journal of Neuroscience Methods*, *193*(1), 156–163. doi: 10.1016/j.jneumeth.2010.08.030

Güler, I., & Übeyli, E. D. (2007). Multiclass support vector machines for eeg-signals classification. *IEEE Transactions on Information Technology in Biomedicine*, *11*(2), 117-126.

Hastie, Trevor Tibshirani, Robert Friedman, J. (2009). *The Elements of Statistical Learning* (Second ed.). doi: 10.1007/b94608

Hussein, R., Mohamed, A., Shaban, K., & Mohamed, A. a. (2013). for Epileptic Detection : A Comparative Study. *Clinical Neurophysiology*, 170–175.

Iscan, Z., Dokur, Z., & Demiralp, T. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Systems with Applications*, *38*(8), 10499–10505. doi: 10.1016/j.eswa.2011.02.110

Jacobs, J., Vogt, C., LeVan, P., Zelmann, R., Gotman, J., & Kobayashi, K. (2016). The identification of distinct high-frequency oscillations during spikes delineates the seizure onset zone better than high-frequency spectral power changes. *Clinical Neurophysiology*, *127*(1), 129–142. doi: 10.1016/j.clinph.2015.04.053

Jensen, O., & Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences*, *11*(7), 267–269. doi: 10.1016/j.tics.2007.05.003

Kubat, M. (2017). *An Introduction to Machine Learning*. Retrieved from `https://link.springer.com/book/10.1007{%}2F978-3-319-63913-0` doi: 10.1007/978-3-319-63913-0

Kwan, P., & Brodie, M. J. (2000, feb). Early Identification of Refractory Epilepsy. *New England Journal of Medicine*, *342*(5), 314–319. doi: 10.1056/NEJM200002033420503

Liu, S., Gurses, C., Sha, Z., Quach, M. M., Sencer, A., Bebek, N., ... Ince, N. F. (2018). Stereotyped high-frequency oscillations discriminate seizure onset zones and critical functional cortex in focal epilepsy. *Brain*, *141*(3), 713–730. doi: 10.1093/brain/awx374

Liu, S., Sha, Z., Sencer, A., Aydoseli, A., Bebek, N., Abosch, A., ... Ince, N. F. (2016). Exploring the time-frequency content of high frequency oscillations for automated identification of seizure onset zone in epilepsy. *Journal of Neural Engineering*, *13*(2). doi: 10.1088/1741-2560/13/2/026026

Malinowska, U., Bergey, G. K., Harezlak, J., & Jouny, C. C. (2015). Identification of seizure onset zone and preictal state based on characteristics of high frequency oscillations. *Clinical Neurophysiology*, *126*(8), 1505–1513. doi: 10.1016/j.clinph.2014.11.007

Malladi, R., Johnson, D. H., Kalamangalam, G. P., Tandon, N., & Aazhang, B. (2018). Data-driven estimation of mutual information using frequency domain and its application to epilepsy. *Conference Record of 51st Asilomar Conference on Signals, Systems and Computers, ACSSC 2017*, *2017-Octob*, 2015–2019. doi: 10.1109/ACSSC.2017.8335721

Matsumoto, A., Brinkmann, B. H., Stead, S. M., Matsumoto, J., Kucewicz, M. T., Marsh, W. R., ... Worrell, G. (2013). Pathological and physiological high-frequency oscillations in focal human epilepsy. *Journal of Neurophysiology*, *110*(8), 1958–1964. doi: 10.1152/jn.00341.2013

Melo, F. (2013). Area under the roc curve. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of systems biology* (pp. 38–39). New York, NY: Springer New York. doi: 10.1007/978-1-4419-9863-7_209

Modur, P., & Miocinovic, S. (2015). Interictal high-frequency oscillations (HFOs) as predictors of high frequency and conventional seizure onset zones. *Epileptic Disorders*, *17*(4), 413–424. doi: 10.1684/epd.2015.0774

Moritz, D. J., Fox, P. J., Luscombe, P. A., & Kraemer, H. C. (1997). Neurological and psychiatric predictors of mortality in patients with Alzheimer disease in California. *Archives of Neurology*, *54*(7), 878–885. doi: 10.1001/archneur.1997.00550190066016

Murphy, P. M., Von Paternos, A. J., & Santaniello, S. (2017). A novel HFO-based method for unsupervised localization of the seizure onset zone in drug-resistant epilepsy. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 1054–1057. doi: 10.1109/EMBC.2017.8037008

Nicolaou, N., & Georgiou, J. (2012). Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines. *Expert Systems with Applications*, *39*(1), 202–209. doi: 10.1016/j.eswa.2011.07.008

Picot, M. C., Baldy-Moulinier, M., Daurès, J. P., Dujols, P., & Crespel, A. (2008). The prevalence of epilepsy and pharmacoresistant epilepsy in adults: A population-based study in a Western European country. *Epilepsia*, *49*(7), 1230–1238. doi: 10.1111/j.1528-1167.2008.01579.x

Price, S. J., & TS. (2008). An algorithm for seizure onset detection using intracranial eeg. *Bone*, *23*(1), 1–7. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf doi: 10.1038/jid.2014.371

Qu, H., & Gotman, J. (1997). A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: Possible use as a warning device. *IEEE Transactions on Biomedical Engineering*, *44*(2), 115–122. doi: 10.1109/10.552241

Saab, M. E., & Gotman, J. (2005). A system to detect the onset of epileptic seizures in scalp EEG. *Clinical Neurophysiology*, *116*(2), 427–442. doi: 10.1016/j.clinph.2004.08.004

Sciaraffa, N., Klados, M. A., Borghini, G., Di Flumeri, G., Babiloni, F., & Aricò, P. (2020). Double-Step Machine Learning Based Procedure for HFOs Detection and Classification. *Brain Sciences*, *10*(4), 220. doi: 10.3390/brainsci10040220

Shoeb, A. (2009). Application of machine learning to epileptic seizure onset detection and treatment. *Diss. Massachusetts Institute of Technology*, 157–162. Retrieved from http://dspace.mit.edu/handle/1721.1/54669

Shoeb, A., Edwards, H., Connolly, J., Bourgeois, B., Ted Treves, S., & Guttag, J. (2004). Patient-specific seizure onset detection. *Epilepsy and Behavior*, *5*(4), 483–498. doi: 10.1016/j.yebeh.2004.05.005

Siuly, S., & Li, Y. (2015). Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. *Computer Methods and Programs in Biomedicine*, *119*(1), 29–42. doi: 10.1016/j.cmpb.2015.01.002

Smart, O., & Chen, M. (2015). Semi-automated patient-specific scalp EEG seizure detection with unsupervised machine learning. *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2015*. doi: 10.1109/CIBCB.2015.7300286

Staba, R. J., Wilson, C. L., Bragin, A., & Fried, I. (2002). Quantitative analysis of high-frequency oscillations (80-500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. *Journal of Neurophysiology*, *88*(4), 1743–1752. doi: 10.1152/jn.2002.88.4.1743

Sumsky, S. L., & Santaniello, S. (2018). Temporal Pattern of Ripple Events in Temporal Lobe Epilepsy: Towards a Pattern-based Localization of the Seizure Onset Zone. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2018-July*, 2288–2291. doi: 10.1109/EMBC.2018.8512742

Tito, M., Cabrerizo, M., Ayala, M., Jayakar, P., & Adjouadi, M. (2010). A comparative study of intracranial EEG files using nonlinear classification methods. *Annals of Biomedical Engineering*, *38*(1), 187–199. doi: 10.1007/s10439-009-9819-6

Turitsyna, E. G., & Webb, S. (2005). Simple design of fbg-based vsb filters for ultra-dense wdm transmission. *Electronics Letters*, *41*(2), 89-91. doi: 10.1049/el:20056760

Van Drongelen, W., Nayak, S., Frim, D. M., Kohrman, M. H., Towle, V. L., Lee, H. C., ... Hecox, K. E. (2003). Seizure anticipation in pediatric epilepsy: Use of kolmogorov entropy. *Pediatric Neurology*, *29*(3), 207–213. doi: 10.1016/S0887-8994(03)00145-0

Varatharajah, Y., Berry, B., Cimbalnik, J., Kremen, V., Van Gompel, J., Stead, M., ... Worrell, G. (2018). Integrating artificial intelligence with real-time intracranial EEG monitoring to automate interictal identification of seizure onset zones in focal epilepsy. *Journal of Neural Engineering*, *15*(4). doi: 10.1088/1741-2552/aac960

Wan, T., Wu, M., Wan, X., & Du, Y. (2016). Automatic detection of high frequency oscillations based on Fuzzy entropy and Fuzzy neural network. *Chinese Control Conference, CCC*, *2016-Augus*, 5027–5032. doi: 10.1109/ChiCC.2016.7554134

Wirth, R., & Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29–39).

Xiang, J., Li, C., Li, H., Cao, R., Wang, B., Han, X., & Chen, J. (2015). The detection of epileptic seizure signals based on fuzzy entropy. *Journal of Neuroscience Methods*, *243*, 18–25. doi: 10.1016/j.jneumeth.2015.01.015

Zhang, Y., Xu, G., Wang, J., & Liang, L. (2010). An automatic patient-specific seizure onset detection method in intracranial EEG based on incremental nonlinear dimensionalityreduction. *Computers in Biology and Medicine*, *40*(11-12), 889–899. doi: 10.1016/j.compbiomed.2010.09.010

Appendix A

Additional content

A.1 Exploratory Data Analysis

A.1.1 Variables Distribution

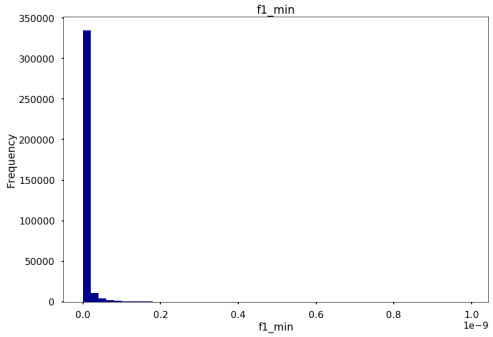

Figure A.1: Histogram - Min

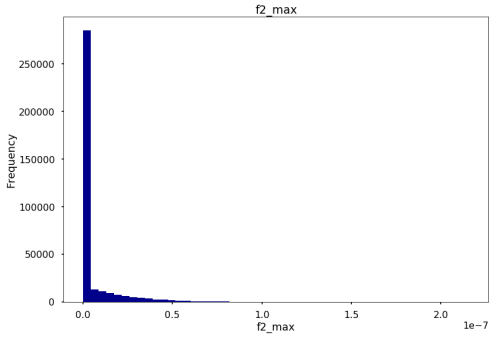

Figure A.2: Histogram - Max

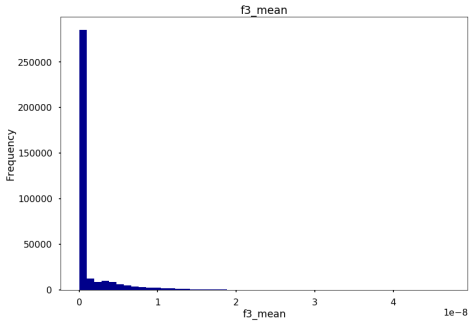

Figure A.3: Histogram - Mean

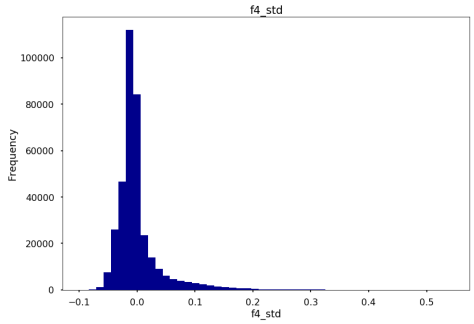

Figure A.4: Histogram - Standard Dev

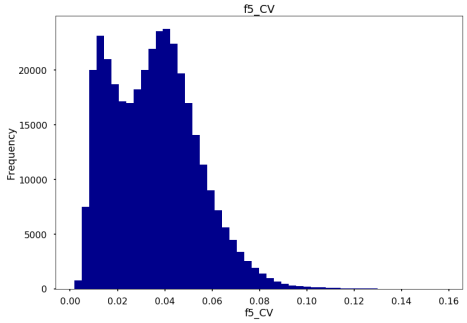

Figure A.5: Histogram - CV

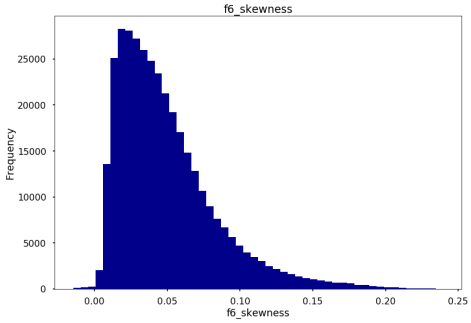

Figure A.6: Histogram - Skewness

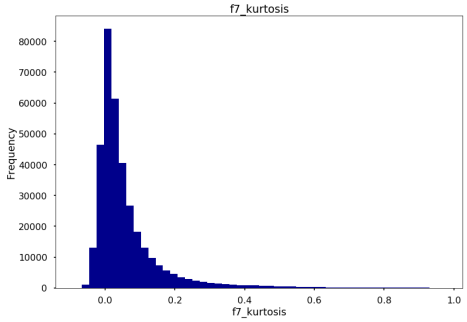

Figure A.7: Histogram - Kurtosis

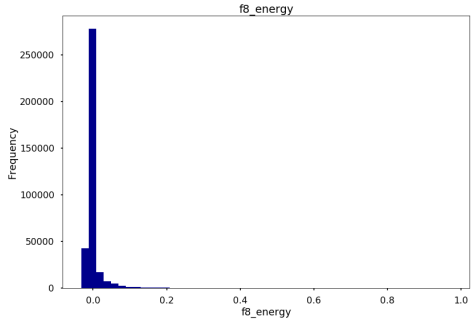

Figure A.8: Histogram - Energy

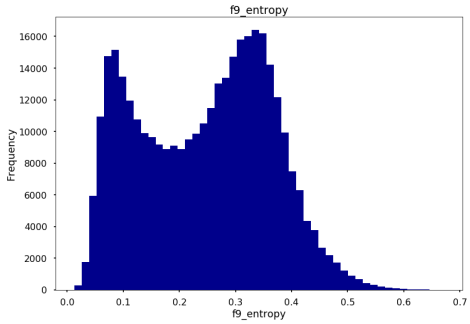

Figure A.9: Histogram - Entropy

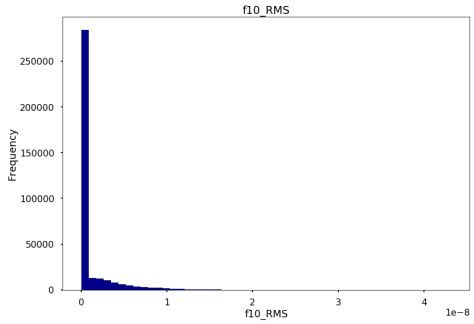

Figure A.10: Histogram - RMS

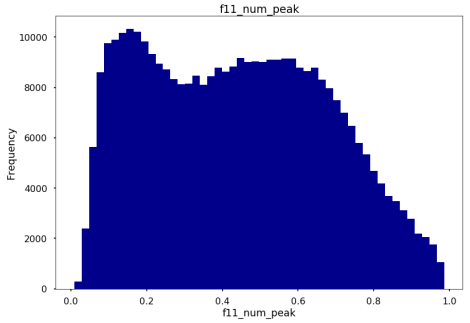

Figure A.11: Histogram - No. of Peaks

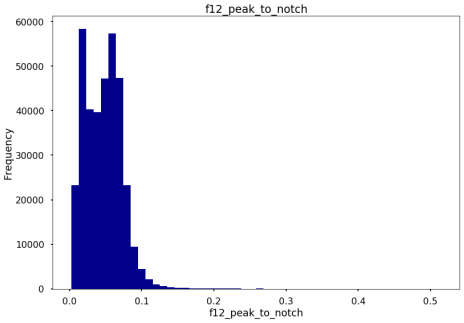

Figure A.12: Histogram - Peak Notch

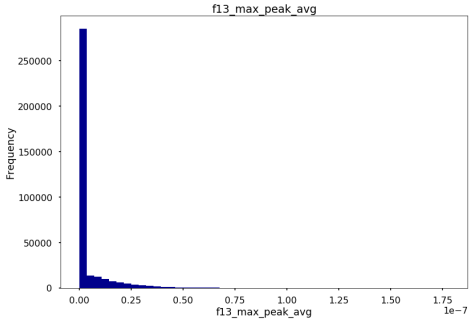

Figure A.13: Histogram - Max peak avg

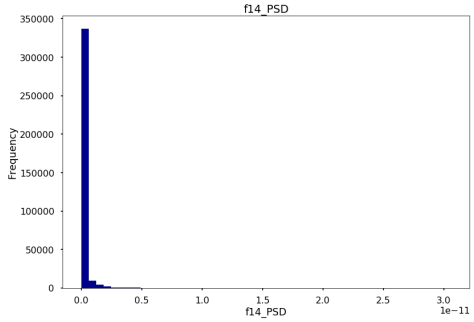

Figure A.14: Histogram - Power Spectral Density

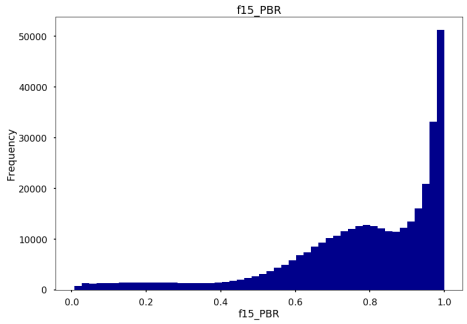

Figure A.15: Histogram - PBR

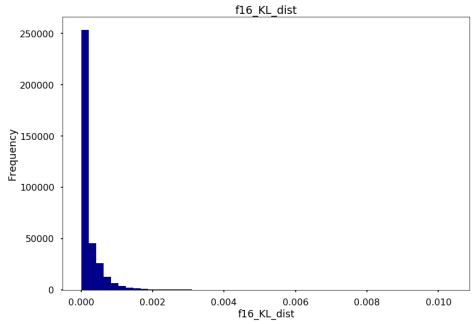

Figure A.16: Histogram - KL dist

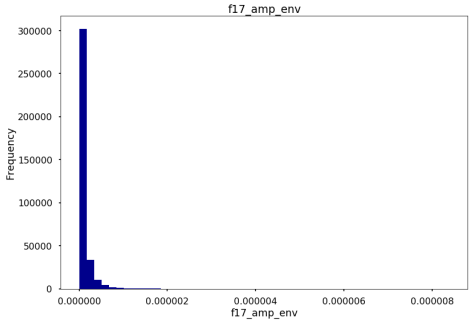

Figure A.17: Histogram - Envelope Amplitude

A.1.2 Variables Correlation

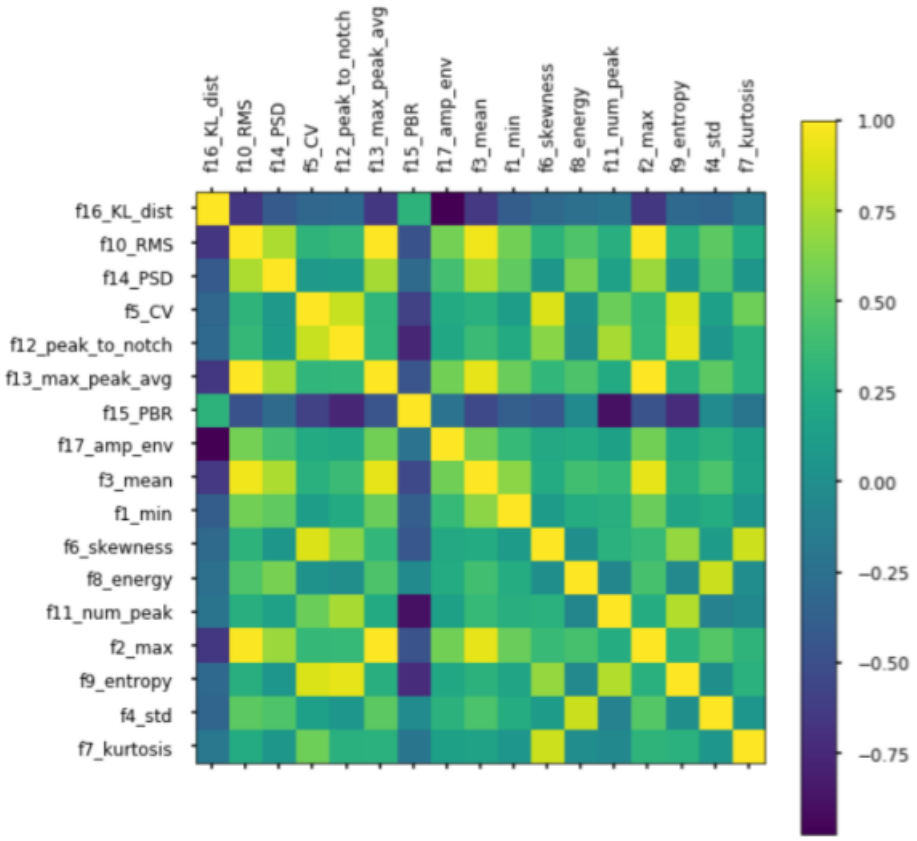

Figure A.18: Correlation Matrix

A.1.3 Bivariate Analysis

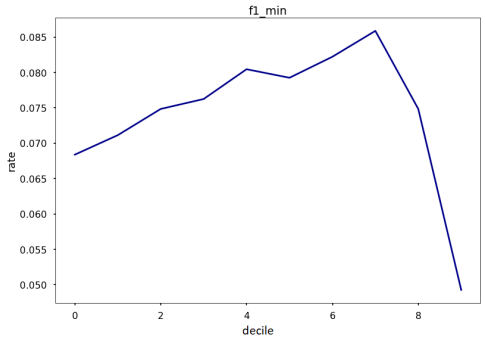

Figure A.19: Bivariate Analysis - Min

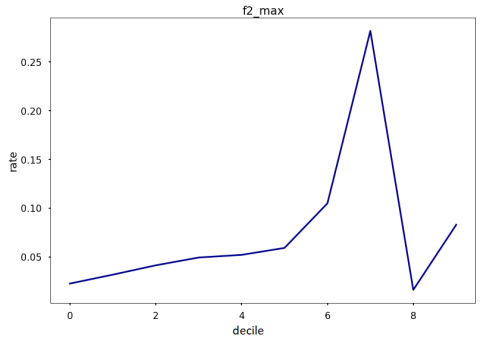

Figure A.20: Bivariate Analysis - Max

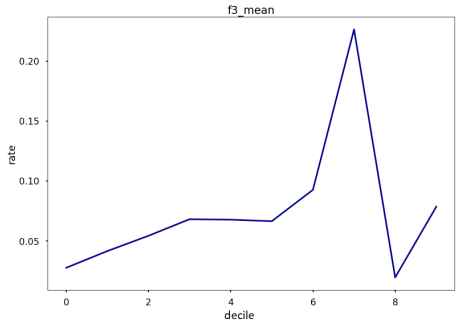

Figure A.21: Bivariate Analysis - Mean

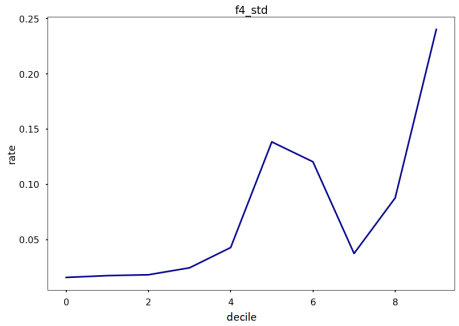

Figure A.22: Bivariate Analysis - Standard Dev

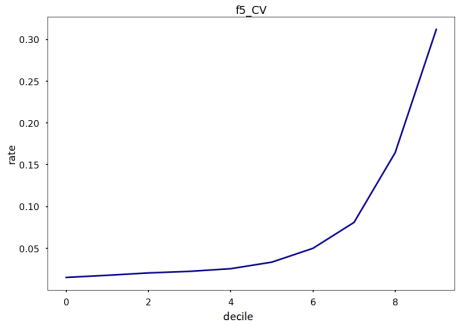

Figure A.23: Bivariate Analysis - Coef. Variation

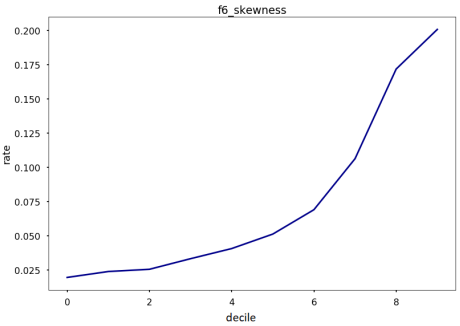

Figure A.24: Bivariate Analysis - Skewness

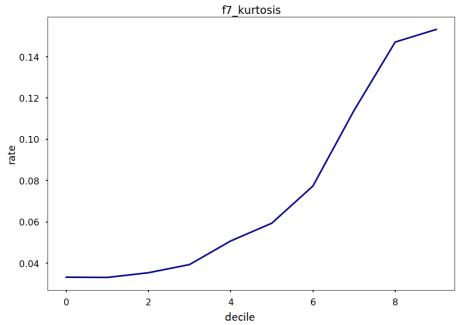

Figure A.25: Bivariate Analysis - Kurtosis

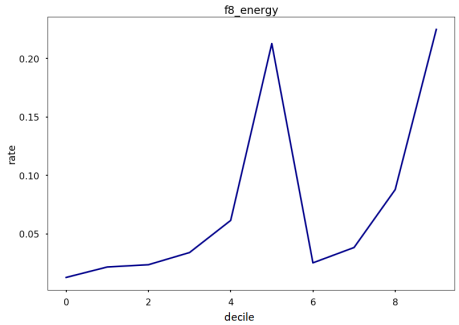

Figure A.26: Bivariate Analysis - Energy

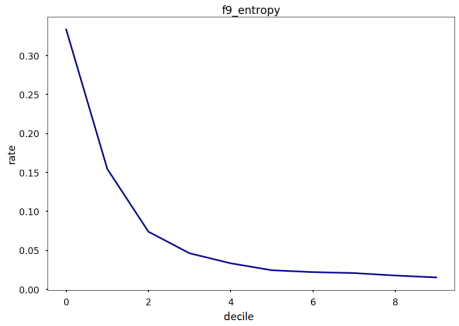

Figure A.27: Bivariate Analysis - Entropy

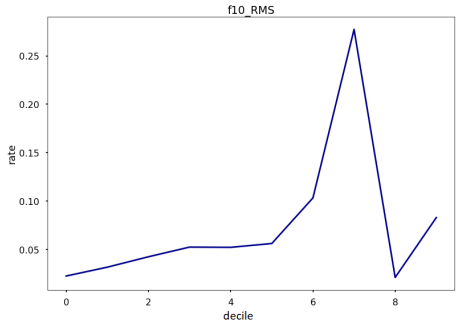

Figure A.28: Bivariate Analysis - Root Mean Sqre

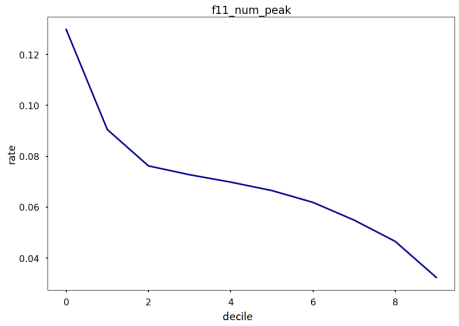

Figure A.29: Bivariate Analysis - No. of Peaks

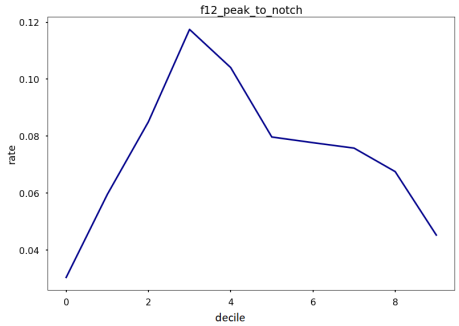

Figure A.30: Bivariate Analysis - Peak Notch

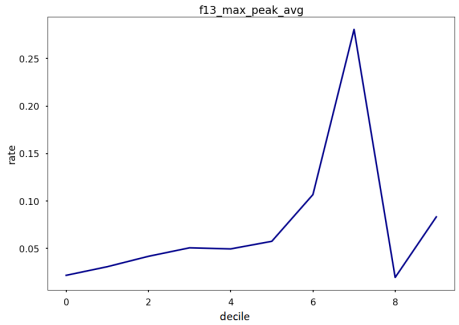

Figure A.31: Bivariate Analysis - Max peak avg

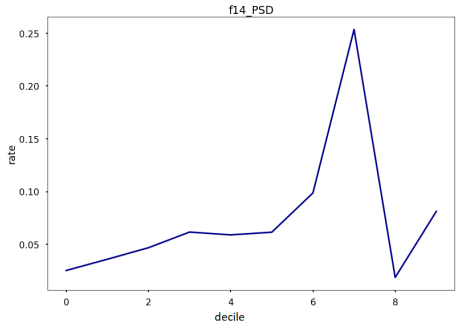

Figure A.32: Bivariate Analysis - Power Spectral Density

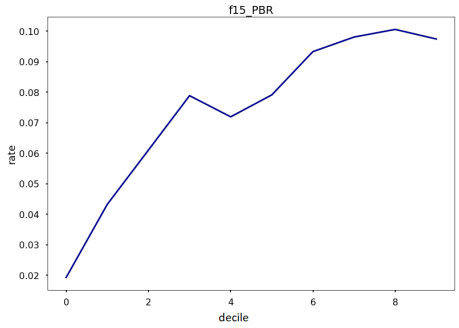

Figure A.33: Bivariate Analysis - Power Band Ratio

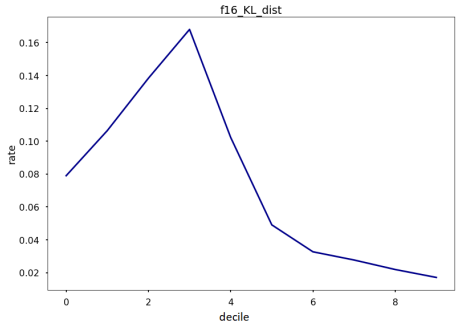

Figure A.34: Bivariate Analysis - KL dist

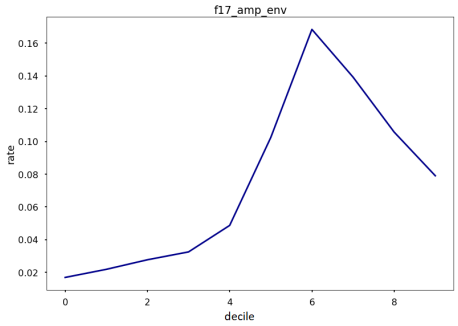

Figure A.35: Bivariate Analysis - Envelope Amplitude